IntechOpen

Telecommunications Networks

Current Status and Future Trends

Edited by Jesús Hamilton Ortiz

TELECOMMUNICATIONS NETWORKS – CURRENT STATUS AND FUTURE TRENDS

Edited by **Jesús Hamilton Ortiz**

Telecommunications Networks - Current Status and Future Trends
http://dx.doi.org/10.5772/2097
Edited by Jesus Hamilton Ortiz

Contributors

Wei Zhuang, Khadija Stewart, James Stewart, Paulo Henrique Carvalho, Marcio De Deus, Priscila Barreto, João Pedro, João Pires, Rafael Marin-Lopez, Fernando Pereniguez-Garcia, Antonio F. Gómez-Skarmeta, Brahim Raouyane, Mostapha Bellafkih, Xiuquan Qiao, Xiaofeng Li, Junliang Chen, Alejandro Muñoz, Luis Zabala, Armando Ferro, Alberto Pineda, Valeriy Bezruk, Matjaž Fras, Joze Mohorko, Zarko Cucej, César Guerra Torres, Jesús De León Morales, Mihael Mohorcic, Ales Svigelj, Andrea Toppan, Paolo Toppan, Cristina De Castro, Oreste Andrisano, Bruno Sericola, Fabrice Guillemin, Dritan Nace, Arta Dilo, Nirvana Meratnia, Ada Gogu, Ciro D'Apice, Benedetto Piccoli, Rosanna Manzo, Carlo Bruni, Francesco Delli Priscoli, Giorgio Koch, Antonio Pietrabissa, Laura Pimpinella, Sergiy O. Gnatyuk, Oleksandr Korchenko, Yevhen Vasiliu, Petro Vorobiyenko, Maksym Lutskiy

© **The Editor(s) and the Author(s) 2012**
The moral rights of the and the author(s) have been asserted.
All rights to the book as a whole are reserved by INTECH. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECH's written permission.
Enquiries concerning the use of the book should be directed to INTECH rights and permissions department (permissions@intechopen.com).
Violations are liable to prosecution under the governing Copyright Law.

Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 3.0 Unported License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be foundat http://www.intechopen.com/copyright-policy.html.

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in Croatia, 2012 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Telecommunications Networks - Current Status and Future Trends
Edited by Jesus Hamilton Ortiz

p. cm.

ISBN 978-953-51-0341-7

eBook (PDF) ISBN 978-953-51-5611-6

We are IntechOpen, the world's leading publisher of Open Access books

Built by scientists, for scientists

4,100+
Open access books available

116,000+
International authors and editors

120M+
Downloads

Our authors are among the

151
Countries delivered to

Top 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Meet the editor

Dr Jesús Hamilton Ortiz has obtained his Bachelors' degrees in Mathematics and Electrical Engineering, DEA in Telecommunications Engineering and a PhD in Computer Engineering. Currently, he is an Assistant Professor at the University of Castilla La Mancha in Computer and Mobile Networks. Professor Ortiz is an editor and reviewer of several international journals and director of closemobile.com. Additionally, he is an assessor of projects concerning applications in Telecommunications and Mobile Networks and a supervisor of bachelor and master degree thesis. He is interested in the following topics: New Generation Networks, 4G, Routing Protocols, QoS, Sensor Networks, VANET, UAVs, etc.

Contents

Preface

In general, all-IP network architecture only provides “Best Effort” services for large volume of data flowing through the network. This massive amount of data and applications in different areas increasingly demand better treatment of the information. Many applications such as medicine, education, telecommunications, natural disasters, stock exchange markets or real-time services, require a superior treatment than the one offered by the “Best Effort” IP protocol.

The new requirements arising from this type of traffic and certain users' habits have produced the necessity of different levels of services and a more scalable architecture, with better support for mobility and increased data security. Large companies are increasing the use of data content, which requires greater bandwidth. Video-conferencing is a good example. There are also delay-sensitive applications like the stock exchange market.

The relentless use of mobile terminals and the growth of traffic over telecommunication networks, whether fixed or mobile, are a true global phenomenon in the field of telecommunications. The increasing use of mobile devices in recent years has been exponential. Nowadays, the number of mobile terminals exceeds that of personal computers. At the same time, we see that mobile networks are a good alternative to complement or replace existing gaps for Internet access in fixed networks, especially in developing countries.

The growth in the use of Telecommunications networks has come mainly with the third generation systems and voice traffic. With the current third generation and the arrival of the 4G, the number of mobile users in the world will exceed the number of landlines users. Audio and video streaming have had a significant increase, parallel to the requirements of bandwidth and quality of service demanded by those applications.

The increase in data traffic is due to the expansion of the Internet and all kinds of data and information on different types of networks. The success of IP-based applications such as web and broadband multimedia contents are a good example. These factors create new opportunities in the evolution of the Telecommunications Networks. Users demand communications services regardless whether the type of access is fixed or via

radio, using mobile terminals. The services that users demand are not only traditional data, but interactive multimedia applications and voice (IMS). To do so, a certain quality of service (QoS) must be guaranteed.

The success of IP-based applications has produced a remarkable evolution of telecommunications into an all-IP network. In theory, the use of IP communications protocol facilitates the design of applications and services regardless the environment where they are used, either a wired or a wireless network. However, IP protocols were originally designed for fixed networks. Their behaviour and throughput are often affected when they are launched over wireless networks.

When it comes to quality of service in communications, IP-based networks alone do not provide adequate guarantees. Therefore, we need mechanisms to ensure the quality of service (QoS) required by applications. These mechanisms were designed for fixed networks and they operate regardless the conditions and status of the network. In wireless networks (Sensor, Manet, etc.), they must be related to the mobility protocols, since the points where a certain quality of service is provided may vary. The challenge is to maintain the requested QoS level while terminals move on and handovers occur.

The technology requires that the applications, algorithms, modelling and protocols that have worked successfully in fixed networks can be used with the same level of quality in mobile scenarios. The new-generation networks must support the IP protocol. This book covers topics key to the development of telecommunications networks researches that have been made by experts in different areas of telecommunications, such as 3G/4G, QoS, Sensor Networks, IMS, Routing, Algorithms and Modelling.

Professor Jesús Hamilton Ortiz
University of Castilla La Mancha
Spain

Part 1

New Generation Networks

1

Access Control Solutions for Next Generation Networks

F. Pereniguez-Garcia, R. Marin-Lopez and A.F. Gomez-Skarmeta
Faculty of Computer Science, University of Murcia
Spain

1. Introduction

In recent years, wireless telecommunications systems have been prevalently motivated by the proliferation of a wide variety of wireless technologies, which use the air as a propagation medium. Additionally, users have been greatly attracted for wireless-based communications since they offer an improved user experience where information can be exchanged while changing the point of connection to the network. This increasing interest has led to the appearance of mobile devices such as smart phones, tablet PCs or netbooks which, equipped with multiple interfaces, allow *mobile users* to access network services and exchange information anywhere and at any time. To support this *always-connected* experience, communications networks are moving towards an *all-IP* scheme where an IP-based network core will act as connection point for a set of accessible networks based on different wireless technologies. This future scenario, referred to as the *Next Generation Networks* (NGNs), enables the convergence of different heterogeneous wireless access networks that combine all the advantages offered by each wireless access technology per se.

In a typical NGN scenario users are expected to be potentially mobile. Equipped with wireless-based multi-interface lightweight devices, users will go about their daily life (which implies to perform movements and changes of location) while demanding access to network services such as VoIP or video streaming. The concept of *mobility* demands session continuity when the user is moving across different networks. In other words, active communications need to be maintained without disruption (or limited breakdown) when the user changes its connection point to the network during the so-called *handoff*.

This aspect is of vital importance in the context of NGNs to allow the user to roam seamlessly between different networks without experiencing temporal interruption or significant delays in active communications. Nevertheless, during the handoff, the connection to the network may for various reasons be interrupted, which causes a packet loss that finally impacts on the on-going communications.

Thus, to achieve mobility without interruptions and improve the quality of the service perceived by the user, it is crucial to reduce the time required to complete the handoff. The handoff process requires the execution of several tasks (N. Nasser et al. (2006)) that negatively affect the handoff latency. In particular, the authentication and key distribution processes have been proven to be one of the most critical components since they require considerable time (A. Dutta et al. (2008); Badra et al. (2007); C. Politis et al. (2004); Marin-Lopez et al. (2010); R. M. Lopez et al. (2007)). The implantation of these processes during the *network access control*

demanded by network operators is destined to ensure that only allowed users can access the network resources in a secure manner. Thus, while necessary, these security services must be carefully taken into account, since they may significantly affect the achievement of seamless mobility in NGNs.

In this chapter we are going to revise the different approaches that have been proposed to address this challenging issue in future NGNs. More precisely, we are going to carry out this analysis in the context of the *Extensible Authentication Protocol* (EAP), a protocol which is acquiring an important position for implementing the access control solution in future NGNs. This interest is motivated by the important features offered by the protocol such as flexibility and media independence. Nevertheless, the EAP authentication process has shown certain inefficiency in mobile scenarios. In particular, a typical EAP authentication involves a considerable signalling to be completed. The research community has addressed this problem by defining the so-called *fast re-authentication* solutions aimed at reducing the latency introduced by the EAP authentication. Throughout this chapter, we will revise the different groups of fast re-authentication solutions according to the strategy followed to minimize the authentication time.

The remaining of the chapter is organized as follows. Section 2 describes the different technologies related to the network access authentication. Next, Section 3 outlines the deficiencies of EAP in mobile environments, which have motivated the research community the proposal of fast re-authentication solutions. The different fast re-authentication schemes proposed so far are analyzed in Section 4. Finally, the chapter finalizes with Section 5 where the most relevant conclusions are extracted.

2. Protocols involved in the network access service

2.1 AAA infrastructures: Authentication, Authorization and Accounting (AAA)

Network operators need to control their subscribers so that only authenticated and authorized ones can access to the network services. Typically, the correct support of a controlled access to the network service has been guaranteed by the deployment of the so-called *Authentication, Authorization and Accounting* (AAA) infrastructures (C. de Laat et al. (2000)). AAA essentially defines a framework for coordinating these individual security services across multiple network technologies and platforms.

An overview of the different components is the best way to understand the services provided by the AAA framework.

- *Authentication*. This service provides a means of identifying a user that requires access to some service (e.g., network access). During the authentication process, users provide a set of credentials (e.g., password or certificates) in order to verify they are who they claim to be. Only when the credentials are correctly verified by the AAA server, the user is granted access to the service.
- *Authorization*. Authorization typically follows the authentication and entails the process of determining whether the client is allowed to perform and request certain tasks or operations. Authorization is the process of enforcing policies, determining what types or qualities of activities, resources or services a user is permitted.
- *Accounting*. The third component in the AAA framework is accounting, which measures the resources a user consumes during network access. This can include the amount of time

a service is used or the amount of data a user has sent and/or received during a session. Accounting is carried out by gathering session statistics and usage information, and it is used for different purposes like billing.

The following sections provide a detailed description for the general AAA architecture and the most relevant AAA protocols.

2.1.1 Generic AAA architecture

The general AAA scheme, as defined in (C. de Laat et al. (2000)), requires the participation of four different entities (see Fig. 1) that take part in the authentication, authorization and accounting processes:

- A *user* desiring to access a specific service offered by the network operator.
- A *domain* where the user is registered. This domain, typically referred to as *home domain*, is able to verify the user's identity based on some credentials. Optionally, the home domain not only authenticates but also provides authorization information to the user
- A *service provider* controlling the access to the offered services. The service provider can be implemented by the domain where the user is subscribed to (home domain) or by a different domain in the roaming cases. In the case the service provider is located outside the home domain, the access to the service is provided on condition that an agreement is established between the service provider and the home domain. These bilateral agreements, which may take the form of formal contracts known as *Service Level Agreements* (SLAs), suppose the establishment of a trust relationship between the involved domains that will allow the service provider to authenticate and authorize foreign users coming from another administrative domains.
- A *service provider's service equipment* which will be typically located on a device that belongs to the service provider. For example, in the case of network access service, this role is played by the *Network Access Server* (NAS) like, for example, an 802.11 access point.

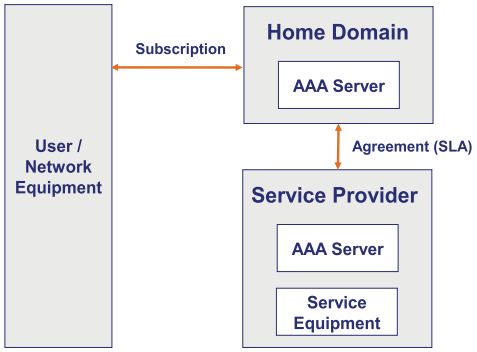


Fig. 1. Generic AAA architecture

2.1.2 Relevant AAA protocols

To allow the communication between AAA servers, it is required the deployment of a *AAA protocol*. Nowadays, the most relevant AAA protocols are RADIUS (C. Rigney et al. (2000)) and Diameter (P. Calhoun & J. Loughney (2003)). Despite Diameter is the most complete AAA protocol, RADIUS is the most widely deployed one in current AAA infrastructures. In the following, it is provided a brief overview of both.

2.1.2.1 RADIUS

RADIUS is a client-server protocol where a NAS usually acts as *RADIUS client*. During authentication procedures, the RADIUS client is responsible for passing user information in the form of requests to the *RADIUS server* and waits for a response from the server. Depending on the policy, the NAS may only need a successful authentication or further authorization directives from the server to enable data traffic to the client. The RADIUS server, on the other hand, is responsible for processing requests, authenticating the users and returning the information necessary for user-specific configuration to deliver the service.

The typical RADIUS conversation consists of the following messages:

- *Access-Request*. This message is sent from the RADIUS client (NAS) to the server to request authentication and authorization for a particular user.
- *Access-Challenge*. This message, sent from the RADIUS server to the client, is used by the server to obtain more information from the NAS about the end user in order to make a decision about the requested service.
- *Access-Accept*. This message is sent from the RADIUS server to the NAS to indicate a successful completion of the request.
- *Access-Reject*. This message is sent by the server to indicate the rejection of a request.

Typically, the main part of a RADIUS conversation consists of several Access-Request/Access-Challenge message exchanges where the RADIUS client and server exchange information transported within RADIUS attributes. Depending on whether the client is successfully authenticated or not, the RADIUS server finalizes the communication with an *Access-Accept* or *Access-Reject*, respectively.

Apart from these main messages, the RADIUS base specification defines some others to transmit accounting information (*Accounting-Request/Accounting-Response*) or the status of the RADIUS entities (*Status-Client/Status-Server*).

Regarding the protocol used to transport RADIUS messages, protocol designers considered that the *User Datagram Protocol* (UDP) was the most appropriate one since the *Transmission Control Protocol* (TCP) session establishment is a time-consuming process requiring the management of connection state. Nevertheless, the lack of a reliable transport causes serious problems to RADIUS. For example, clients are unable to distinguish when a request is received by the server or a communication problem has occurred and the RADIUS packet has not reached its destination. Similarly, a client cannot distinguish whether a server is down or discarding requests.

RADIUS security is another aspect that was not deeply considered. In particular, it is based on the use of shared secrets between the RADIUS client and the server. In real deployments, this basic security mechanism has been known to cause several vulnerabilities:

- Shared secrets must be statically configured. No method for dynamic shared secret establishment is defined in the RADIUS protocol.
- Shared secrets are determined according to the source IP address in the RADIUS packet. This introduces management problems when the client's IP address change.
- When using RADIUS proxies, the RADIUS client only shares a secret with the RADIUS server in the first hop and not with the ultimate RADIUS server. In other words, the trust

relationship between the RADIUS client and the final RADIUS server is transitive rather than using a direct trust relationship. If a server in the chain is compromised, some security problems arise.

- RADIUS does not provide high transport protection. For example, an observer can examine the content of RADIUS messages and trace the content of a specific attribute.

To overcome these security weakness, it has been proposed the use of TLS (T. Dierks & C. Allen (1999)) to provide a means to secure the RADIUS communication between client and server on the transport layer (S. Winter et al. (2010)). Nevertheless, the main research and standardization efforts have focused on the design of a new AAA protocol called *Diameter*.

2.1.2.2 Diameter

Diameter, proposed as an enhancement to RADIUS, is considered the next generation AAA protocol. Diameter is characterized by its extensibility and adaptability since it is designed to perform any kind of operation and supply new needs that may appear in future control access technologies. Another cornerstone of Diameter is the consideration of multi-domain scenarios where AAA infrastructures administered by different domains are interconnected to provide an unified authentication, authorization and accounting framework. For this reason, Diameter is widely used in 3G networks and its adoption is recommended in future AAA infrastructures supporting access control in NGN.

The Diameter protocol defines an extensible architecture that allows to incorporate new features through the design of the so-called *Diameter applications*, which rely on the basic functionality provided by the *base protocol*. The Diameter *base protocol* (P. Calhoun & J. Loughney (2003)), defines the Diameter minimum elements such as the basic set of messages, attribute structure and some essential attribute types. Additionally, the basic specification defines the inter-realm operations by defining the role of different types of Diameter entities. Diameter applications are services, protocols and procedures that use the facilities provided by the Diameter base protocol itself. Every Diameter application defines its own *commands* and *messages* which, in turn, can define new attributes called *Attribute Value Pair* (AVP) or re-use existing ones already defined by some other applications.

The Diameter base protocol does not define any use of the protocol and expects the definition of specific applications using the Diameter functionality. For example, the use of Diameter for providing authentication during network access is defined in the *Diameter NAS Application* (P. Calhoun et al. (2005)). In turn, this specification is used by the *Diameter EAP Application* (P. Eronen et al. (2005)) to specify the procedure to perform the network access authentication by using the EAP protocol. Similarly, authorization and accounting procedures are expected to be handled by specific applications.

Within a Diameter-based infrastructure, the protocol distinguishes different types of nodes where each one plays a specific role:

1. *Diameter Client*: represents an entity implementing network access control like, for example, a NAS. The Diameter client issues messages soliciting authentication, authorization or accounting services for a specific user.
2. *Diameter Server*: is the entity that processes authentication, authorization and accounting request for a particular domain. The Diameter server must support the Diameter base protocol and the applications used in the domain.

3. *Diameter Agent*: is an entity that processes a request and forwards it to a Diameter server or to another agent. Depending on the service provided, we can distinguish:
   (a) *Relay agents*: which forward messages based on routing-related attributes and routing tables.
   (b) *Proxy agents*: which act as a relay agent that, additionally, may modify the routed message based on some policy.
   (c) *Redirect agents*: instead of routing messages, they inform the sender about the proper way to route the message.
   (d) *Translation agents*: which perform protocol translations between Diameter and other AAA protocols such as RADIUS.

The different types of nodes exchange Diameter messages that carry information. Instead of defining a message type, Diameter uses the concept of *command* to specify the type of function a Diameter message intends to perform. Because the message exchange style of Diameter is synchronous, each command consists of a request and its corresponding answer. Table 1 provides a brief summary of the main Diameter commands defined in the base protocol specification.

Command	**Abbreviation**	**Description**
Capabilities-Exchange- Request /Answer	CER/CEA	Discovery of a peer's identity and its capabilities.
Disconnect-Peer-Request /Answer	DPR/DPA	Used to inform the intention of shutting down the connection.
Re-Auth-Request /Answer	RAR/RAA	Sent to an access device (NAS) to solicit user re-authentication.
Session-Termination-Request /Answer	STR/STA	To notify that the provision of a service to a user has finalized.
Accounting-Request /Answer	ACR/ACA	To exchange accounting information between Diameter client and server.

Table 1. Common Diameter commands

2.2 The Extensible Authentication Protocol (EAP)

The *Extensible Authentication Protocol* (EAP) (B. Aboba et al. (2004)) is a protocol designed by the *Internet Engineering Task Force* (IETF) that permits the use of different types of authentication mechanisms through the so-called *EAP methods* (e.g., based on symmetric keys, digital certificates, etc.). These are performed between an *EAP peer* and an *EAP server*, through an *EAP authenticator* which merely forwards EAP packets back and forth between the EAP peer and the EAP server. From a security standpoint, the EAP authenticator does not take part in the mutual authentication process but acts as a mere EAP packet forwarder.

One of the advantages of the EAP architecture is its flexibility since does not impose a specific authentication mechanism. Additionally, EAP is independent of the underlying wireless access technology, being able to operate in NGNs. Finally, EAP allows an easy integration with existing Authentication, Authorization and Accounting (AAA) infrastructures (B. Aboba et al. (2008) by defining a configuration mode that permits the use of a backend authentication server, which may implement some authentication methods. These advantages have motivated the success of the EAP authentication protocol for network access control in future NGNs.

2.2.1 Components

The EAP protocol consists of request and response messages. Request messages are sent from the authenticator to the peer. Conversely, response messages are sent from the peer to the authenticator. The different messages exchanged during an EAP execution are processed by several components that are conceptually organized in four layers:

- *EAP Lower-Layer*. This layer is responsible for transmitting and receiving EAP packets between the peer and authenticator.
- *EAP Layer*. The EAP layer is responsible for receiving and transmitting EAP packets through the transport layer. The EAP layer not only forwards packets between the EAP transport and peer/authenticator layers, but also implements duplicate detection and packet retransmission.
- *EAP Peer / Authenticator Layer*. EAP assumes that an EAP implementation will support both the EAP peer and the authenticator functionalities. For this reason, based on the code of the EAP packet, the EAP layer demultiplexes incoming EAP packets to the EAP peer and authenticator layers.
- *EAP Method Layer*. An EAP method implements a specific authentication algorithm that requires the transmission of EAP messages between peer and authenticator.

2.2.2 Distribution of the EAP entities

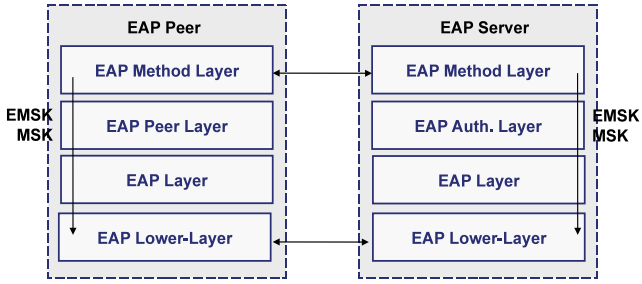
As previously mentioned, an EAP authentication involves three entities: the EAP peer, authenticator and server. Whereas the EAP peer is co-located with the mobile, the EAP authenticator is commonly placed on the *Network Access Server* (NAS) (e.g., an access point or an access router). Depending on the location of the EAP server, two authenticator models have been defined. Figures 2(a) and 2(b) show the *standalone authenticator model* and the *pass-through authenticator model*, respectively. On the one hand, in the standalone authenticator model (Fig. 2(a)), the EAP server is implemented on the EAP authenticator. On the other hand, in the pass-through authenticator model (Fig. 2(b)), the EAP server and the EAP authenticator are implemented in separate nodes.

In order to deliver EAP messages, an *EAP lower-layer* (e.g., IEEE 802.11) is used to transport the EAP packets between the EAP peer and the EAP authenticator. The protocol used to transport messages between the EAP authenticator and the EAP server depends on the authenticator model employed. More precisely, in the standalone authenticator model, the communication between the EAP server and standalone authenticator occurs locally in the same node. In the pass-trough authenticator model, the EAP protocol requires help of an auxiliary AAA protocol such as RADIUS or Diameter.

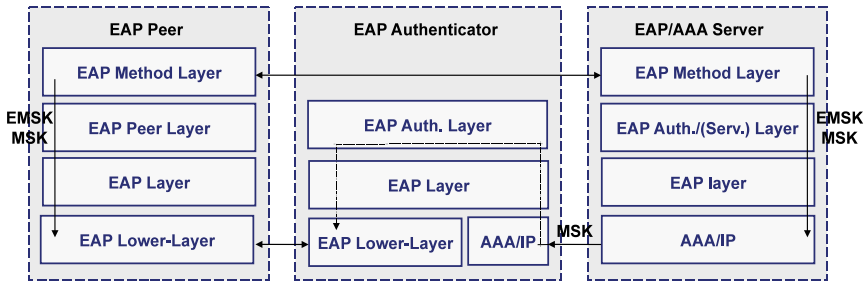
2.2.3 EAP authentication phases

As depicted in Fig. 3, a typical EAP conversation [1] occurs in three different phases. Initially, in the discovery phase (*Phase 0*), the peer discovers the EAP authenticator near to the peer's location with which it desires to start an authentication process. This phase, which is supported by the specific EAP lower-layer protocol, can be performed either manually or automatically.

[1] Without loss of generality, it is assumed an EAP pass-through authenticator model.



(a) Standalone Authenticator Model



(b) Pass-through Authenticator Model

Fig. 2. EAP authenticator models

The authentication phase (*phase 1*) starts when the peer decides to initiate an authentication process with a specific authenticator. This phase consists of two steps. Firstly, the *phase 1a* includes an EAP authentication exchange between the EAP peer, authenticator and server. To start an EAP authentication, the EAP authenticator usually starts the process by requesting the EAP peer's identity through an *EAP Request/Identity* message. The trigger that signals the EAP authenticator to start the EAP authentication is outside the scope of EAP. Examples of these triggers are the *EAPOL-Start* message defined in IEEE 802.1X (IEEE 802.11 (2007)) or simply an 802.11 association process. On the reception of the *EAP Request/Identity*, the EAP peer answers with an *EAP Response/Identity* with its identity. With this information, the EAP server will select the EAP method to be performed. The EAP method execution involves several exchanges of *EAP Request* and *EAP Response* messages between the EAP server and the EAP peer. A successful EAP authentication finishes with an *EAP Success* message.

Certain EAP methods (Dantu et al. (2007)) are able to generate key material. In particular, according to the *EAP Key Management Framework* (EAP KMF) (B. Aboba et al. (2008)) two keys are exported after a successful EAP authentication: the *Master Session Key* (MSK) and the *Extended Master Session Key* (EMSK). The former is traditionally sent (using the AAA protocol) to the authenticator (*Phase 1b*) to establish a security association with the EAP peer (*Phase 2*). Instead, the latter must not be provided to any other entity outside the EAP server and peer. Thus, both entities may use the EMSK for further key derivation. In particular, as we will analyze in Section 4, some authentication schemes propose to employ the EMSK to derive further key material for enabling a fast re-authentication process.

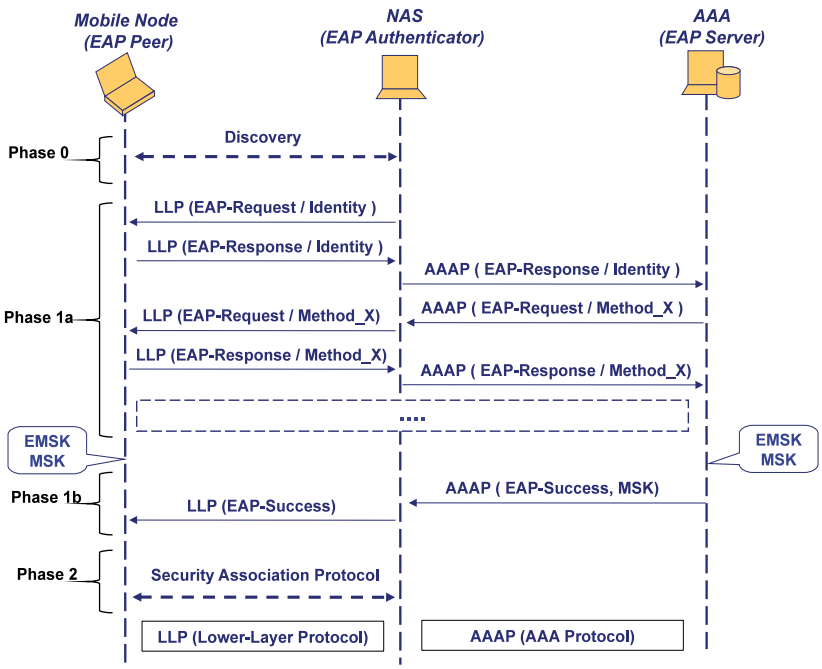


Fig. 3. EAP authentication exchange

2.3 Existing technologies for network access control

The EAP lower-layer protocol allows an EAP peer to perform an EAP authentication process with an authenticator. Basically, the EAP lower-layer is responsible for transmitting and receiving EAP packets between peer and authenticator. Currently, a wide variety of lower-layer protocols can be found since each link-layer technology defines its own transport to carry EAP messages (e.g., IEEE 802.1X, IEEE 802.11, IEEE 802.16e). However, there are also lower-layer protocols operating at network level which are able to transport EAP messages on top of IP (e.g., PANA). Finally, some other lower-layer protocols provide an hybrid solution to transport EAP packets either at link-layer or network layer (e.g., IEEE 802.21 MIH). In the following, the most representative technologies for network access control are analyzed.

2.3.1 IEEE 802.1X

The IEEE 802.1X specification (IEEE 802.1X (2004)) is an access control model developed by the *Institute of Electrical and Electronics Engineers* (IEEE) that allows to employ different authentication mechanisms by means of EAP in IEEE 802 *Local Area Networks* (LANs). As depicted in Fig. 4, there are three main components in the IEEE 802.1X authentication system: *supplicant*, *authenticator* and *authentication server*. In a *Wireless LAN* (WLAN), the supplicant is usually a mobile user, the access point usually represents an authenticator and an AAA server is the authentication server. 802.1X defines a mechanism for port-based network access control. A port is a point through which a supplicant can access to a service offered by a device. The port in 802.1X represents the association between the supplicant and the authenticator. Both the supplicant and the authenticator have a PAE (*Port Access Entity*) that operates the algorithms and protocols associated with the authentication process.

Initially, as depicted in Fig. 4, the authenticator's controlled port is in unauthorized state, that is, the port is *open*. Only received authentication messages will be directed to the authenticator PAE, which will forward them to the authentication server. This initial configuration allows to unauthenticated supplicants to communicate with the authentication server in order to perform an authentication process based on EAP. Once the user is successfully authenticated, the PAE will close the controlled port, allowing the supplicant to access the network service offered by the authenticator's system.

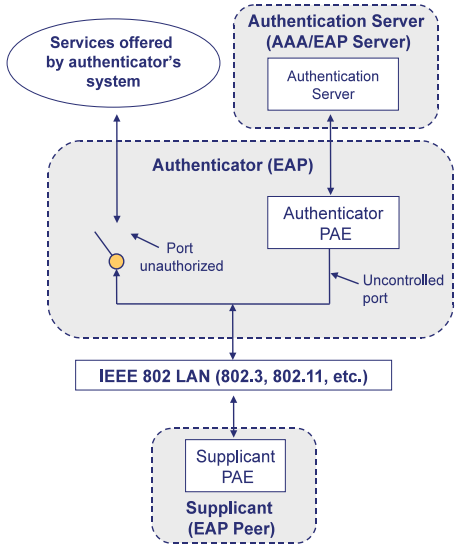


Fig. 4. IEEE 802.1X architecture

2.3.2 IEEE 802.11

IEEE 802.11 extends the IEEE 802.1X access control model by defining algorithms and protocols to protect the data traffic between *station* (STA) and *access point* (AP). More precisely, once the EAP authentication is successfully completed, both STA and AP will share a *Pairwise Master Key* (PMK). This key, derived from the MSK exported by the EAP authentication, is used by a security association protocol (called *4-way handshake*) intended to negotiate cryptographic keys to protect the wireless link between STA and AP. Once the security association is successfully established, the controlled port is closed and access to the network is granted to the supplicant.

The authentication process, described in Fig. 5, involves three entities: an STA acting as supplicant, an AP acting as authenticator and an authentication server (e.g., an AAA server) that assists the authentication process. The process starts with the so-called *IEEE 802.11 association phase* where the STA firstly discovers the security capabilities implemented by the AP *(1)*. Next, the IEEE 802.11 authentication exchange *(2)* is invoked in order to maintain backward compatibility with the IEEE 802.11 state machine. This exchange is followed by an association process *(3)* where the negotiation of the cryptographic suite used to protect the traffic is performed.

In the subsequent *IEEE 802.11 authentication phase*, an EAP authentication is performed where the STA acts as *EAP peer* and the AP acts as *EAP authenticator (4)*. Conversely, the *EAP*

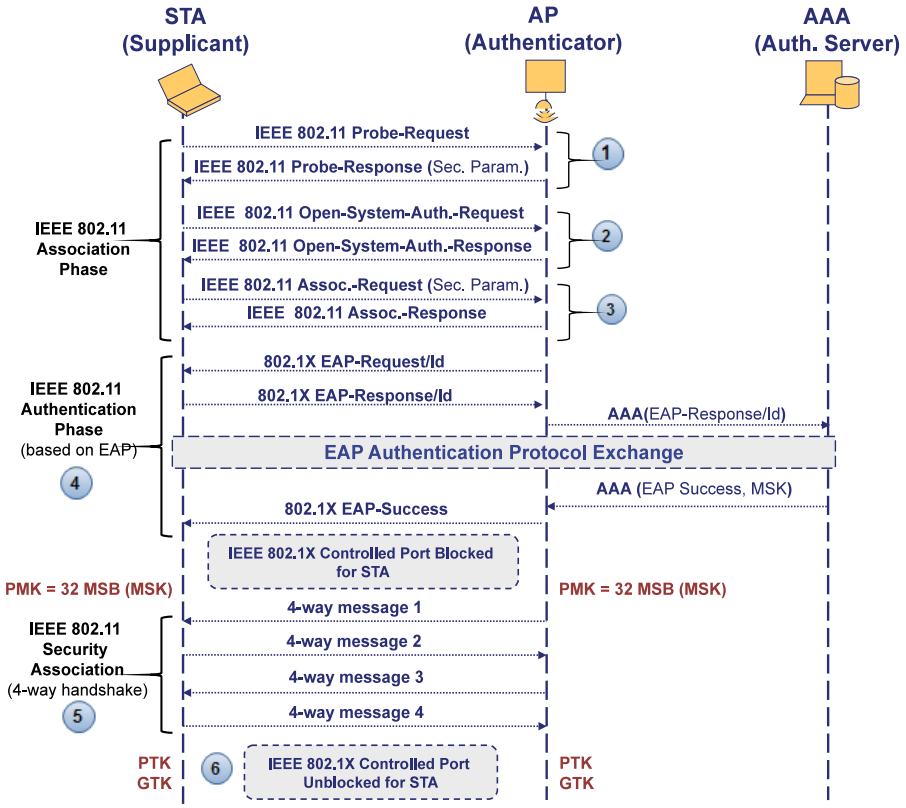


Fig. 5. IEEE 802.11 message flow

server can be co-located with the EAP authenticator (*standalone configuration*) or within an external authentication server (*pass-through configuration*), in which case an AAA protocol (e.g., RADIUS or Diameter) is used to transport EAP messages between the authenticator and the server. Once the EAP authentication is successfully completed, the 32 *more significant bytes* (MSB) from the exported MSK is used as PMK.

Following the establishment of the PMK, a *4-way handshake* protocol is executed during the *IEEE 802.11 security association phase (5)* to confirm the existence of the PMK and selected cryptographic suites. The protocol generates a *Pairwise Transient Key* (PTK) for unicast traffic and a *Group Transient Key* (GTK) for multicast traffic. Thus, as result of a successful *4-way handshake*, a secure communication channel between the STA and the AP is established for protecting data traffic in the wireless link.

2.3.3 IEEE 802.16e

The IEEE 802.16e (*IEEE 802.16e* (2006)) specification is an extension for IEEE 802.16 networks that enables the mobility support and enhances the basic access control mechanism defined for fixed scenarios in order to provide authentication and confidentiality in IEEE 802.16-based wireless networks. In particular, the security architecture is further strengthened by introducing the *Privacy and Key Management* protocol version 2 (PKMv2) which provides mutual authentication and secure distribution of key material between the IEEE 802.16

subscriber station (SS) and the *base station* (BS). The authentication can be performed by using an EAP-based authentication scheme.

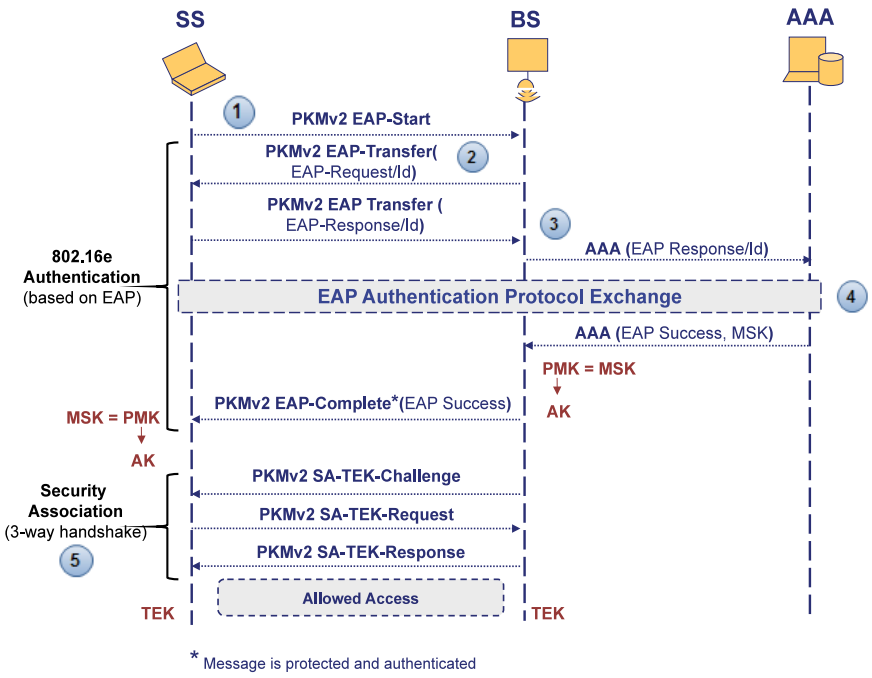


Fig. 6. IEEE 802.16e message flow

Figure 6 shows the authentication process. As observed, while the SS acts as *EAP peer*, the BS implements the *EAP authenticator* functionality. Depending on the EAP configuration mode, the *EAP server* can be placed in the BS (*standalone mode*) or in a AAA server (*pass-through*), which is the case assumed in Fig. 6. As observed, while EAP messages exchanged between SS and BS are transported within the *PKMv2 EAP-Transfer* message, an AAA protocol (e.g., RADIUS or Diameter) is used to convey EAP messages between the BS and the AAA server.

Once the EAP authentication is successfully completed, from the exported MSK a *Pairwise Master Key* (PMK) is derived. In turn, from this PMK, an *Authorization Key* (AK) is generated for the security association establishment. For this reason, the 802.16e specification requires the use of EAP methods exporting key material. Finally, as previously mentioned, the AK shared between SS and BS is employed by a security association protocol called *3-way handshake (5)*, which verifies the possesion of the AK and generates a *Traffic Encryption Key* (TEK) used to protect the traffic in the wireless link.

2.3.4 PANA

The *Protocol for carrying Authentication for Network Access* (PANA) (D. Forsberg et al. (2008)) is a network-layer transport for authentication information designed by the IETF *PANA Working Group* (PANA WG). PANA is designed to carry EAP over UDP to support a variety of authentication mechanisms for network access (thanks to EAP) as well as a variety of underlying network access technologies (thanks to the use of UDP). As highlighted in Fig. 7, PANA considers a network access control model integrated by the following entities:

- The *PANA Client* (PaC) is the client implementation of PANA. This entity resides on the subscriber's node which is requesting network access. The PaC acts as EAP peer according to the EAP model described earlier.
- The *PANA Authentication Agent* (PAA) is the server implementation of PANA. A PAA is in charge of communicating with the PaCs for authenticating and authorizing them to access the network service. The PAA acts as EAP authenticator.
- The *Enforcement Point* (EP) refers to the entity in the access network in charge of inspecting data traffic of authenticated and authorized subscribers. Basically, the EP represents a point of attachment (e.g., access point) to the network.
- The *Authentication Server* (AS) is in charge of verifying the credentials provided by a PaC through a PAA. The AS functionality is typically implemented by an AAA server, which also integrates the EAP server.

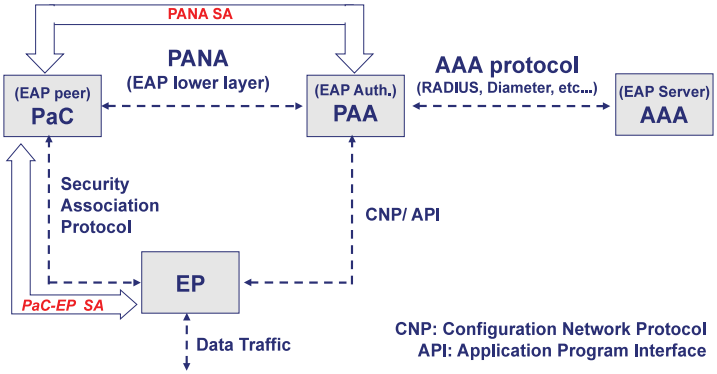


Fig. 7. PANA architecture

Additionally, there are two types of security associations related to PaC in the PANA architecture. On the one hand, a *PANA security association* (PANA SA) is established between the PaC and PAA in order to integrity protect PANA messages. On the other hand, a *PaC-EP SA* is established by performing a security association protocol between the PaC and an EP to protect data traffic.

The PANA operation is developed along four different phases. Initially, during the *authentication and authorization phase*, the PaC and the PAA negotiate some parameters, such as the integrity algorithms used to protect PANA messages. They also exchange PANA messages transporting EAP to perform the authentication and establish a so-called *PANA session*. If the PaC is successfully authenticated, the protocol enters in the *access phase* where the PaC can use the network service by just sending data traffic through the EP. If the PANA session is about to expire, typically a *re-authentication phase* happens to renew this session lifetime. Finally, the PaC or PAA can terminate the session (e.g., the PaC desires to log out the network access session) during *termination phase*, where resources allocated by the network for the PaC are also removed. If neither PaC nor PAA can complete the termination phase, both entities can release the resources once the PANA session lifetime expires.

During each phase, a different set of messages can be sent. Basically we can find four types of PANA messages.

- *PANA-Client-Initiation* (PCI). This message is sent by the PaC requesting the PAA start the authentication process.

- *PANA-Auth-Request/Response* (PAR/PAN). These messages are used during the authentication and authorization phase and the re-authentication phase. They allow to negotiate some parameters between the PaC and the PAA and to carry authentication information in the format of EAP packets.
- *PANA-Notification-Request/Response* (PNR/PNA). These messages are exchanged once PaC is authenticated. They are used as keep-alive mechanism of the PANA authentication session or to signal the beginning of a re-authentication process.
- *PANA-Termination-Request/Response* (PTR/PTA). These messages are used to end up a PANA session.

2.3.5 IEEE 802.21 MIH

The IEEE 802.21 is a recent effort that aims at enabling seamless service continuity among heterogeneous networks (IEEE 802.21 (2008); Taniuchi et al. (2009)). The standard defines a logical entity, *MIH Function* (MIHF), which facilitates the mobility management and handover process. The MIHF is located within the mobility management protocol stack of a mobile node (MN) or network entity. Through the media independent interface, MIHF supports useful services (events, commands or information) that help in determining the need for initiate a handoff or selecting a candidate network

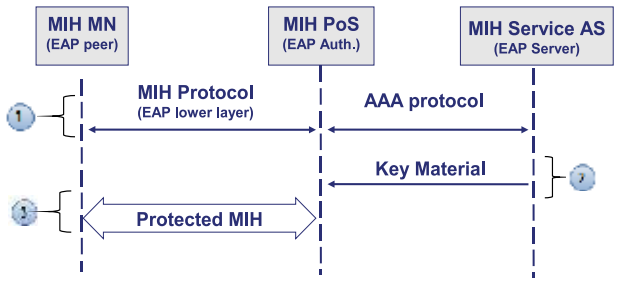


Fig. 8. MIH protocol as EAP lower-layer

Different *tasks groups* (TG) have defined extensions to IEEE 802.21. For example, the standardization task group IEEE 802.21a is defining mechanisms that allow to protect the IEEE 802.21 MIH protocol messages. The solution (EAP over MIH (2010)) designed by the task group proposes that the *mobile node* (MN) must be authenticated and authorized before granting access to the services offered by the *Point of Service* (PoS). In particular, EAP has been proposed as one alternative to carry out this authentication process. Figure 8 depicts the general process followed to perform an EAP-based *Media-Independent Authentication Process*. As observed, the MN and PoS acts as EAP peer and authenticator, respectively. The EAP server functionality is implemented by an entity named *Service Authentication Server* (Service AS). Initially, an EAP authentication (*1*) is performed between the MN and the Service AS through the PoS, which acts as authenticator. While the MIH protocol is used as EAP lower-layer to transport EAP messages between MN and PoS, an AAA protocol is employed between PoS and Service AS for the same purpose. Note that, since MIH protocol is independent from the underlying transport, this is an hybrid solution that can operate either at link-layer or network-layer. When the EAP authentication is completed, the Service AS sends the MSK (*2*) exported by the EAP method to the PoS. From this MSK, a key hierarchy is generated to protect MIH protocol packets (*3*).

3. Fast re-authentication to optimize the network access control

As we can observe, EAP is a promising authentication protocol to be used in NGNs due to its flexibility, wireless technology independence and integration with AAA infrastructures. Furthermore, it is used by a wide variety of network access technologies as standard solution for authentication. However, EAP has shown some drawbacks when mobility is taken into consideration. The reason why the EAP authentication process is not so optimized for mobile scenarios is due to two main motives. First, a typical EAP authentication requires several message exchanges between EAP peer and server. Depending on the EAP method in use (R. Dantu et al. (2007)), this number can vary. For example, one of the most common methods, EAP-TLS (D. Simon et al. (2008)), involves in the best case up to eight messages between peer and server to complete. Secondly, each round-trip is performed with the EAP server placed on the EAP peer's home domain, where the peer is subscribed to. Especially in roaming scenarios, the EAP server may be far from the mobile user (EAP peer) and, therefore, the latency introduced per each exchange increases. These issues are raised when an EAP peer moves from one authenticator to another (*inter-authenticator handoff*). In this case, the peer needs to perform an EAP authentication with the EAP server, through the new EAP authenticator. Therefore, every time the EAP peer moves to a new EAP authenticator, it may suffer from high handoff latency during EAP authentication.

This problem can affect the on-going communications since the latency introduced by the EAP authentication during the handoff process may provoke a substantial packet loss, resulting in a degradation in the service quality perceived by the user. In this sense, the performance requirements of a real-time application will vary according to the type of application and its characteristics such as delay and packet-loss tolerance. The ITU-T G.114 recommendation (ITU-T Recommendation G.114 (1998)) indicates, for Voice over IP applications, an end-to-end delay of 150 ms as the upper limit and rates 400 ms as a generally unacceptable delay. Similarly, a streaming application has tolerable packet-error rates ranging from 0.1 to 0.00001 with a transfer delay of less than 300 ms. As has been proved in (R. M. Lopez et al. (2007)), a full EAP authentication[2] based on a typical EAP method such as EAP-TLS can provoke an unacceptable handoff interruption of about 600 milliseconds (or even in some cases several seconds) for these kind of applications.

To solve this problem, it is necessary to define a *fast re-authentication process* (T. Clancy et al. (2008)) to reduce the authentication time required by a user to complete an EAP-based authentication. Researchers have not ignored this challenging aspect and a wide set of fast re-authentication mechanisms can be found in the literature. Before analyzing the different fast re-authentication schemes in next Section 4, we are going to present both the desired design and security goals that a proper fast re-authentication mechanism should accomplish. To be aware of these requirements is useful to determine advantages and disadvantages when analyzing the different fast re-authentication solutions.

3.1 Design goals

A suitable fast re-authentication solution should accomplish the following requirements and aims (T. Clancy et al. (2008)):

[2] Note that the term *full* is used in comparison with *reduced* to denote that, in the execution of an EAP method, there is no optimization to reduce the number of exchanges during the EAP authentication.

(D1) *Low latency operation*. The fast re-authentication mechanism must reduce the authentication time executed during the network access control process compared with a traditional full EAP authentication. Furthermore, the achievement of a reduced handoff latency must not affect the security of the authentication process.

(D2) *EAP lower-layer independence*. Any keying hierarchy and protocol defined must be independent of the lower-layer protocol used to transport EAP packets between the peer and the authenticator. In other words, the fast re-authentication solution must be able to operate over heterogeneous technologies, which is the expected scenario in NGNs. Nevertheless, in certain circumstances, the fast re-authentication mechanism could require some assistance from the lower layer protocol.

(D3) *Compatibility with existing EAP methods*. The adoption of a fast re-authentication solution must not require modifications to existing EAP methods. In the same manner, additional requirements must not be imposed on future EAP methods. Nevertheless, the fast re-authentication solution can enforce the employment of EAP methods following the *EAP Key Management Framework* (B. Aboba et al. (2008)).

(D4) *AAA protocol compatibility and keying*. Any modification to the EAP protocol itself or the key distribution scheme defined by EAP, must be compatible with currently deployed AAA protocols. Extensions to both RADIUS and Diameter to support these EAP modifications are acceptable. However, the fast re-authentication solution must satisfy the requirements for the key management in AAA environments (B. Aboba et al. (2008); R. Housley & B. Aboba (2007)).

(D5) *Compatibility with other optimizations*. The fast re-authentication solution must be compatible with other optimizations destined to reduce the handoff latency already defined by other standards.

(D6) *Backward compatibility*. The system should be designed in such a manner that a user not supporting fast re-authentication should still function in a network supporting fast re-authentication. Similarly, a peer supporting fast re-authentication should still operate in a network not supporting the fast re-authentication optimization.

(D7) *Low deployment impact*. In order to support the aforementioned design goals, a fast re-authentication solution may require modifications in EAP peers, authenticators and servers. Nevertheless, in order to favour the protocol deployment, the required changes must be minimized (ideally, they should be avoided) in current standardized protocols and technologies.

(D8) *Support of different types of handoffs*. The fast re-authentication mechanism must be able to operate in any kind of handoff regardless of whether it implies a change of technology (intra/inter-technology), network (intra/inter-network), administrative domain (intra/inter-domain) or type of security required by the authenticator (intra/inter-security).

3.2 Security goals

In addition to the aforementioned design goals, a secure fast re-authentication mechanism should accomplish the following security goals (R. Housley & B. Aboba (2007)):

(S1) *Authentication*. This requirement mandates that a management and key distribution mechanism must be designed to allow all parties involved in the protocol execution to authenticate every entity with which it is communicating. That is, it must be feasible to

gain assurance that the identity of the another entity is as declared, thereby preventing impersonation. To carry out the authentication process, it is necessary to define the so-called *security associations* between the involved entities.

(S2) *Authorization*. During the network access control process, the user is not only authenticated but also authorized to access the network service. The authorization decision is taken by the AAA server and the result is communicated to the authenticator. The fast re-authentication solution proposed must not hinder the authorization process performed once the user is successfully authenticated.

(S3) *Key context*. This requirement establishes that any key must have a well-defined scope and must be used in a specific context for an intended use (e.g., cipher data, sign, etc.). During the time a key is valid, all the entities that are authorized to have access to the key must share the same key context. In this sense, keys should be uniquely named so that they can be identified and managed effectively. Additionally, it must be taken into account that the existence of a hierarchical key structure imposes some additional restrictions. For example, the lifetime of lower-level keys must not exceed the lifetime of higher-level keys.

(S4) *Key freshness*. A key is fresh (from the viewpoint of one party) if it can be guaranteed to be recent and not an old key being reused for malicious actions by either an attacker or unauthorized party (A. Menezes et al. (1996)). Mechanisms for refreshing keys must be provided within the re-authentication solution.

(S5) *Domino effect*. In network security, the compromise of keys in a specific level must not result in compromise of other keys at the same level or higher levels that were used to derive the lower-level keys. Assuming that each authenticator is distributed a key to carry out the fast re-authentication process, a key management solution respecting this property will be resilient against the *domino effect* (R. Housley & B. Aboba (2007)) attack, so the compromise of one authenticator must not reveal keys in another authenticators.

(S6) *Transport aspects*. The solution developed must be independent of any underlying transport protocol. Depending on the physical architecture and the functionality of the involved entities, there may be a need for multiple protocols to perform the transport of keying material between entities involved in the fast re-authentication architecture. As far as possible, protocols already designed and used should be used to address the cryptographic material distribution. For example, while AAA protocols can be considered for this purpose between the EAP authenticator and server, the EAP protocol can be used between EAP peer and server.

4. Overview of existing fast re-authentication schemes

This section analyzes the different efforts that have attempted to reduce the EAP authentication time during the network access control process. According to the strategy followed to achieve this objective, the different fast re-authentication solutions can be classified in different groups: *context transfer*, *pre-authentication*, *key pre-distribution*, *use of a local server* and *modifications to EAP*. In the following, we delve into each of them and detail the mechanism proposed to achieve a reduced handoff latency.

4.1 Context transfer

As depicted in Fig. 9, the context transfer mechanism (T. Aura & M. Roe (2005), H. Kim et al. (2005), C. Politis et al. (2004), *IEEE 802.11 IAPP* (2003), J. Bournelle et al. (2006)) tries

to reduce the time devoted to network access control by transferring cryptographic material (*1*) from an EAP authenticator (*current*) to a new one (*target*). When the user moves to the new authenticator (*2*), it can use the transferred context (e.g., cryptographic keys and associated lifetimes) to execute a security association protocol with the new authenticator (*3*) to protect the wireless link. Thus, the user does not need to be authenticated and can directly start the security association establishment process based on the transferred cryptographic material.

In order to perform a secure transference between both authenticators, it is assumed the existence of a pre-established security association between them. Additionally, context transfer solutions do not propagate the same cryptographic material (*CM*) from one authenticator to another. Instead, the transferred cryptographic material is derived (*CM'*) from that owned by the current authenticator where the user is connected. The process employed to generate the derived cryptographic material is followed by both the peer and the authenticator. While the authenticator transfers the derived material to the new authenticator, the peer employs it to start the security protocol execution.

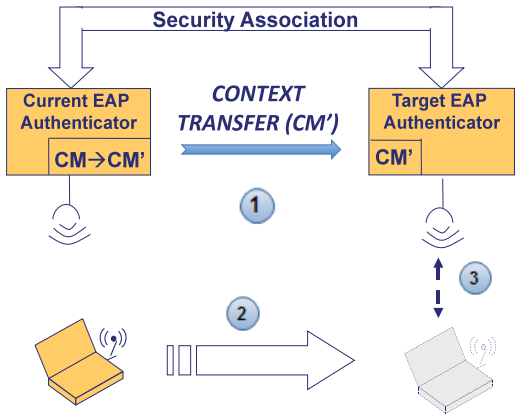


Fig. 9. Context transfer mechanism

Depending on when the transference is performed, we can distinguish between *reactive* and *proactive* schemes. In the proactive mode, the context transfer is performed before the peer performs the handoff. Therefore, when the peer moves to the new authenticator, the cryptographic material has been already transferred to the new authenticator and the peer can immediately establish the security association. Conversely, in the reactive mode, the context transfer is performed once the user performs the handoff and is under the coverage area of the new authenticator. The proactive mode introduces less latency to network access control than the reactive mode since the transference of cryptographic material is performed in advance before the handoff. Nevertheless, reactive solutions are interesting in situations where the handoff happens unexpectedly and there is no anticipation to perform the transference.

An important advantage of context transfer mechanisms relies on their ability to re-authenticate the user without the need of contacting an authentication server located in the infrastructure. Nevertheless, they have been widely criticized as a promising technique to achieve a fast network access due to an important security vulnerability known as the *domino effect* (R. Housley & B. Aboba (2007)). The problem comes from the fact that context transfer re-uses the same cryptographic material (or a derived one following a well-known process) in different authenticators. Therefore, if one authenticator is compromised, the rest of authenticators visited by the same user are also affected.

4.2 Pre-authentication

Pre-authentication solutions propose a scheme (see Fig. 10) where the mobile user performs a full EAP authentication (*1*) with a candidate authenticator through the current associated one *before* it performs the handoff. In this manner, when the handoff happens (*2*), given that the MSK generated during the pre-authentication process will be already present in the candidate authenticator, the peer only needs to establish a security association (*3*) with it to protect the wireless link. As we see, pre-authentication decouples the authentication and network access control operations from the handoff.

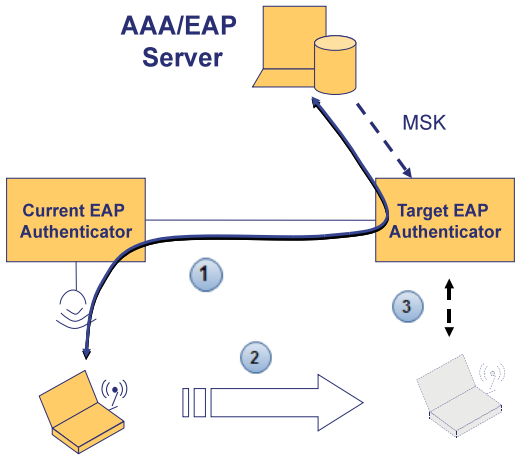


Fig. 10. Pre-authentication mechanism

Depending on the role adopted by the current authenticator during the EAP pre-authentication, we can distinguish two scenarios of EAP pre-authentication signalling (Y. Ohba et al. (2010)):

- *Direct pre-authentication*. In this type of EAP pre-authentication, the current authenticator only forwards the EAP lower-layer messages between mobile node and candidate authenticator as it would be data traffic.
- *Indirect pre-authentication*. Here, the current authenticator plays an active role during pre-authentication process. This type of pre-authentication is useful when the mobile node neither has the candidate authenticator address nor is able to access to the candidate authenticator for security reasons. Therefore, there is a signalling from mobile node to/from current authenticator, and from/to the current authenticator to/from the candidate authenticator. Note that current authenticator does not act as an EAP authenticator; it only translates between different EAP lower-layer protocols.

The first pre-authentication proposal was initially introduced at link layer by the IEEE 802.11i technology (IEEE 802.11i (2005)) and later improved in IEEE 802.11r (IEEE 802.11r (2005)). Nevertheless, the definition of pre-authentication mechanisms at link-layer has some serious limitations since they cannot be applied for cases involving inter-domain or inter-technology handoffs. To avoid this problems, some other solutions propose a pre-authentication procedure at network layer. Network layer solutions (Y. Ohba and A. Yegin (2010), R. M. Lopez et al. (2007), A. Dutta et al. (2008)) have the advantage of being capable to work independent of the underlying access technologies and with authenticators located in different networks or domains.

Despite pre-authentication solutions can potentially achieve an important reduction in the latency introduced by the authentication process during the network access control, this technique presents some drawbacks. First, pre-authentication requires the existence of network connectivity to carry out the pre-authentication process which is a requisite that may not always be satisfied. Second, pre-authentication requires a precise selection of the authenticator with which perform a pre-authentication process. If the user performs a pre-authentication with authenticators where the user finally does not move, the technique may incur in an unnecessary use of network resources. The third disadvantage is related to the previous one. Since pre-authentication implies the pre-reservation of resources in candidate authenticators, in practice, operators are reluctant to pre-reserve resources for users that may or may not roam in the future. Therefore, pre-authentication may have a limited application, specially in inter-domain handoffs. Finally, given that pre-authentication involves a full EAP authentication, special care must be taken to determine the moment to start the pre-authentication process. As a consequence, pre-authentication needs to be performed with a considerable anticipation to the handoff.

4.3 Key pre-distribution

Key pre-distribution solutions (A. Mishra et al. (2004), S. Pack & Y. Choi (2002), Z. Cao et al. (2011), F.Bernal-Hidalgo et al. (2011)) propose the pre-installation of cryptographic material (e.g., keys) in candidate authenticators so that the keys required for secure association are already available when the peer moves to the authenticators. As depicted in Fig. 11, the mobile user initially performs an EAP authentication (*1*) with the AAA server. Once the EAP authentication is successfully completed, the AAA server pre-distributes keys (*2*) to authenticators which the user can potentially associate to in a near future. Therefore, when the peer moves to a new authenticator (*3* and *5*), it is not required to perform a full EAP authentication. Instead, using the key material already present in the authenticator and known by the peer, a security association is established between both entities (*4* and *6*).

Fast re-authentication solutions based on key pre-distribution have two main disadvantages. On the one hand, they require a precise selection of those authenticators to which pre-distribute key material. If the user pre-distributes key material to authenticators where the user finally does not move, the technique may incur in an unnecessary use of resources. Nevertheless, this is a complex problem given the difficulty of predicting future movements of the user. On the other hand, key pre-installation solutions have a significant deployment cost since a modification in existing lower-layer technologies and AAA protocols is required in order to allow pushing a key provided by an external entity instead of being produced as a consequence of a successful EAP authentication executed through the EAP authenticator.

4.4 Use of a local server

According to the EAP authentication model (B. Aboba et al. (2004)), each time a user needs to be authenticated, a full EAP authentication must be performed with the AAA/EAP server located in the user's home domain. This is a serious limitation for roaming scenarios, specially in mobility contexts. The reason is that each time the visited network needs to re-authenticate the client, the home domain must be contacted. This introduces a considerable latency during network access process since the home EAP server could be located far from the current user's location. Furthermore, taking into account that typical EAP methods (e.g., EAP-TLS) require multiple round trips, the home domain needs to be contacted several times in order to complete the EAP conversation, resulting in unacceptable handoff times.

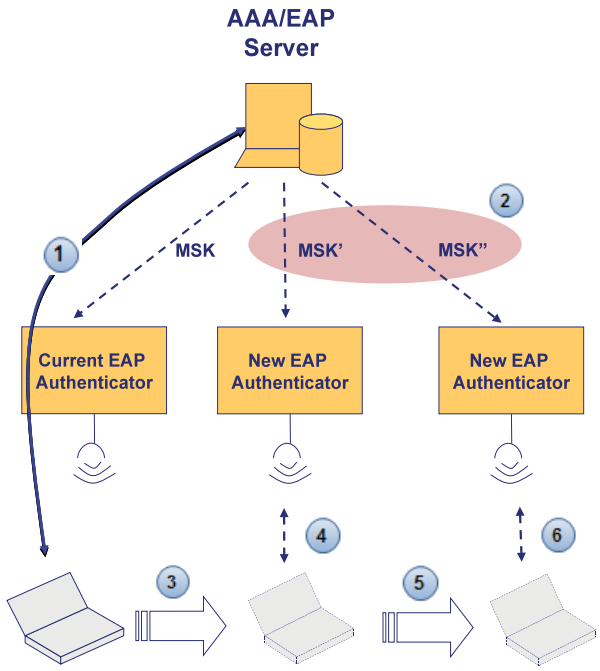


Fig. 11. Key pre-distribution mechanism

To solve this issue, some solutions (3GPP TS 33.102 V7.1.0 (2006), R. Marin et al. (2006), F.Bernal-Hidalgo et al. (2011), V. Narayanan & L. Dondeti (2008)) have proposed the use of a local server near the area of movement of the peer to speed up the re-authentication. The basic idea is to allow the visited domain to play a more active role in network access control by allowing the home AAA server to delegate the re-authentication task to the local AAA server placed in the visited domain. As depicted in Fig. 12, the user firstly performs a full EAP authentication (*1*) with the home AAA/EAP server using the *long-term* credentials that the home domain provides to their subscribers. This initial EAP authentication, commonly named *bootstrapping phase*, is performed the first time the user connects to the network. Next, once the EAP authentication is successfully completed, the home AAA/EAP server sends (*2*) some key material (KM) to the visited AAA/EAP server. This key material, which is used as *mid-term* credential between the mobile and the visited AAA/EAP server, allows to locally perform re-authentication (*3, 4*) when the peer moves to other authenticators located in the visited domain, thus avoiding AAA signalling with the home AAA/EAP server.

Despite this kind of fast re-authentication solutions do not require to contact the home domain to re-authenticate the user, they do not define any optimization for the re-authentication process with the local server. For example, authors in (R. Marin et al. (2006)) propose the use of an EAP method based on shared secret key like EAP-GPSK which requires two message exchanges with the local authentication server. Another serious disadvantage is found in the process followed to distribute the key that establishes a trust relationship between the peer and the local server. Solutions like (F.Bernal-Hidalgo et al. (2011); R. Marin et al. (2006)) use a two-party model to carry out a key distribution process which involves three entities: peer, local re-authentication server and home AAA/EAP server. Since the use of a two-party model is known to be inappropriate (D. Harskin et al. (2007)) from a security standpoint, a three-party approach is recommended.

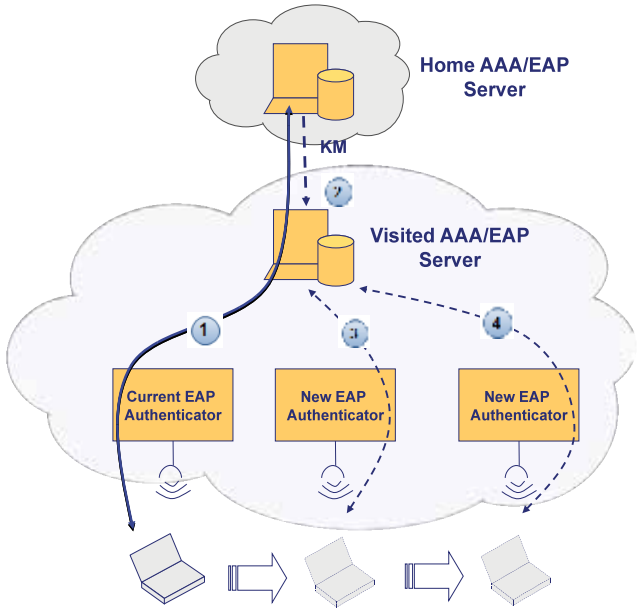


Fig. 12. Use of a local server mechanism

4.5 Modifications to EAP

Finally, another group of solutions try to reduce the EAP authentication time by modifying the EAP protocol itself. Between the different solutions following this approach, the most relevant contribution is the *EAP Extensions for EAP Re-authentication Protocol* (ERP) (V. Narayanan & L. Dondeti (2008)), which has been proposed by the IETF *HandOver KEYing Working Group* (HOKEY WG).

ERP is a method-independent solution that modifies the EAP protocol to achieve a lightweight authentication process. Additionally, ERP relies on the local server optimization (see Section 4.4) and assumes the existence of a local *EAP Re-authentication* (ER) server to optimize the process, which will be in charge of both fast EAP re-authentication and key distribution tasks. The ERP protocol describes a set of extensions to EAP in order to enable efficient re-authentication for a peer that has already established some EAP key material with the EAP server in a previous *bootstrapping phase*. These extensions include three new messages: *EAP-Initiate/Re-auth-Start*, *EAP-Initiate/Re-auth* and *EAP-Finish/Re-auth*.

As shown in Fig. 13, the ERP negotiation involves the peer, the authenticator and the ER server. Beforehand, it is assumed that the peer performs a full EAP authentication with the ER server and both entities share a EMSK. From the EMSK, the peer and the ER server derives a key named rRK. In turn, from the rRK, a new key named *Re-authentication Integrity Key* (rIK) is derived to provide proof of possession and authentication during the re-authentication process.

The ERP re-authentication process is initiated by the authenticator by sending *EAP-Initiate/Re-auth-Start* to the peer. On the reception of this message, the peer sends an *EAP-Initiate/Re-auth* protected with the rIK which is forwarded by the authenticator to the ER server. Once the ER server successfully verifies this messages, it

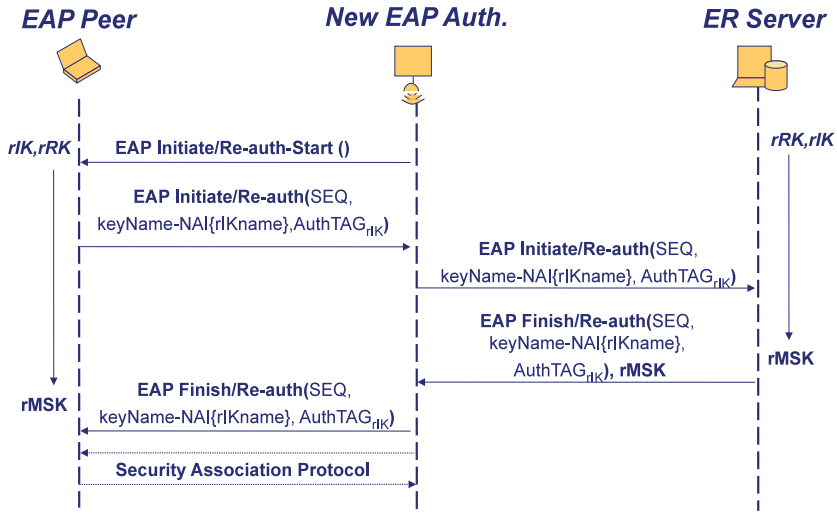


Fig. 13. ERP protocol

replays with a final *EAP-Finish/Re-auth* and derives a rMSK (from the rRK), which is sent to the authenticator to establish a security association with the peer.

On the one hand, in general, the main problem of this kind of proposals relies on their high deployment cost. Since these solutions update the EAP protocol basic operation, they require the modification of existing EAP implementations in order to support the new re-authentication functionality. Consequently, user equipments, authenticators and authentication servers need to be updated, thus complicating the adoption of the solution. On the other hand, in particular, an important drawback of ERP is found on the security of the re-authentication process. Similarly to solutions (F.Bernal-Hidalgo et al. (2011); R. Marin et al. (2006)) previously analyzed in Section 4.4, ERP follows an inappropriate two-party key distribution model to distribute the rMSK from the ER to the authenticator.

5. Conclusion

The provision of *seamless mobility* has created an interesting research field within NGNs in order to find mechanisms which try to provide a continuous access to the network during the handoff. In fact, this is a critical process, where the connection to the network is interrupted, thus causing packet loss that may affect on-going communications. To solve this problem, efforts are directed at reducing the time required to complete the different tasks performed during the handoff. In particular, the network access control process has been demonstrated to be one of the most important factors that negatively affects handoff latency. This process is demanded by network operators in order to control that only legitimate users are able to employ the operator's resources.

This chapter has provided a general overview about the state-of-art of technologies and protocols related to network access control in future NGNs. In particular, we have reviewed the EAP/AAA framework as a promising architecture for network access authentication in future heterogeneous networks. While AAA infrastructures provide an unified framework to handle the authentication, authorization and accounting processes, the EAP protocol is used to implement the authentication service in AAA scenarios. Apart from being easily

deployable within existing AAA infrastructures, EAP exhibits important features such as flexibility to select an authentication mechanism and independence from the underlying wireless technology.

Nevertheless, EAP presents some deficiencies when applied in mobile scenarios. In particular, a typical EAP authentication introduces a prohibitive latency during the handoff which provokes a connection disruption that may affect active communications. This problem has been extensively studied by the research community, which has proposed different fast re-authentication mechanisms.

Precisely, the second part of the chapter is devoted to revise and analyze the different schemes that have tried to reduce the latency introduced by network access control during the handoff. According to the strategy followed to reduce the authentication time, we can distinguish five fast re-authentication schemes: context transfer, pre-authentication, key pre-distribution, use of a local server and modifications to EAP. Throughout this chapter we have analyzed both advantages and disadvantages of each approximation.

6. Acknowledgements

This work is partially supported by the Funding Program for Research Groups of Excellence (04552/GERM/06) and the Spanish Ministry of Science and Education (TIN2008-06441-C02-02).

7. References

3GPP TS 33.102 V7.1.0 (2006). 3rd Generation Partnership Project.

A. Dutta, D. Famolari, S. Das, Y. Ohba, V. Fajardo, K. Taniuchi, R. Lopez & H. Schulzrinne (2008). *Media-Independent Pre-Authentication Suppporting Secure Interdomain Handover Optimization*, *IEEE Wireless Communications* vol. 15(2): 55–64.

A. Menezes, P. van Oorschot & S. Vanstone (1996). Handbook of Applied Cryptography, CRC Press.

A. Mishra, M. Shin, N. Petroni, C. Clancy & W. Arbaugh (2004). *Proactive Key Distribution Using Neighbor Graphs*, *IEEE Wireless Communication* 11: 26–36.

B. Aboba, D. Simon & P. Eronen (2008). *Extensible Authentication Protocol Key Management Framework*. RFC 5247.

B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson & H. Levkowetz (2004). *Extensible Authentication Protocol (EAP)*. RFC3748.

Badra, M., Urien, P. & Hajjeh, I. (2007). *Flexible and fast security solution for wireless LAN*, *Pervasive and Mobile Computing Journal* 3: 1–14.

C. de Laat, G. Gross, L. Gommans, J. Vollbrecht & D. Spence (2000). *Generic AAA Architecture*. IETF RFC 2903.

C. Politis, K. Chew, N. Akhtar, M. Georgiades, R. Tafazolli & T. Dagiuklas (2004). *Hybrid multilayer mobility management with AAA context transfer capabilities for all-IP networks*, *IEEE Wireless Communications 11* pp. pp. 76–88.

C. Rigney, S. Willens, A. Rubens & W. Simpson (2000). *Remote Authentication Dial In User Service (RADIUS)*. IETF RFC 2865.

D. Forsberg, Y. Ohba, B. Patil, H. Tschofenig & A. Yegin (2008). *Protocol for Carrying Authentication for Network Access (PANA)*. IETF RFC 5191.

D. Harskin, Y. Ohba, M. Nakhjiri & R. Marin (2007). *Problem Statement and Requirements on a 3-Party Key Distribution Protocol for Handover Keying*. IETF Internet Draft, draft-ohba-hokey-3party-keydist-ps-01.

D. Simon, B. Aboba & R. Hurst (2008). *The EAP-TLS Authentication Protocol*. IETF RFC 5216.

Dantu, R., Clothier, G. & Atri, A. (2007). EAP Methods for Wireless Networks, *Computer Standards Interfaces* 29(3): 289–301.

EAP over MIH (2010). Option III: EAP to conduct service authentication and MIH packet protection (21-10-0078-08-0sec-option-iii-eap-over-mih-service-authentication).

F.Bernal-Hidalgo, Marin-Lopez, R. & Gomez-Skarmeta, A. (2011). *Key Distribution Mechanisms For IEEE 802.21-Assisted Wireless Heterogeneous Networks*, *Mobile Networks and Management*, Vol. 68, Springer Berlin Heidelberg, pp. 123–134.

H. Kim, K. G. Shin & W. Dabbous (2005). *Improving Cross-domain Authentication over Wireless Local Area Networks*, *Proc. of 1st International Conference on Security and Privacy for Emerging Areas in Communications Networks, SECURECOMM'05*, IEEE Computer Society, Athens, Greece, pp. 103–109.

IEEE 802.11 (2007). Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Network – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.

IEEE 802.11i (2005). Std., Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Specification for Enhanced Security.

IEEE 802.11 IAPP (2003). IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation.

IEEE 802.11r (2005). , Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment 8: Fast BSS Transition.

IEEE 802.16e (2006). Air Interface for Fixed and Mobile Broandband Wireless Access System.

IEEE 802.1X (2004). Standards for Local and Metropolitan Area Networks: Port based Network Access Control, IEEE Standards for Information Technology.

IEEE 802.21 (2008). Institute of Electrical and Electronics Engineers, Draft IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services.

ITU-T Recommendation G.114 (1998). ITU-T General Characteristics of International Telephone Connections and International Telephone Circuits: One-Way Transmission Time, ITU-T Recommendation G.114.

J. Bournelle, M. Laurent-Maknavicius, H. Tschofenig, Y. El Mghazli, G. Giaretta, R. Lopez & Y. Ohba (2006). *Use of Context Transfer Protocol (CXTP) for PANA*. IETF Internet Draft, draft-ietf-pana-cxtp-01.

Marin-Lopez, R., Pereniguez, F., Bernal, F. & Gomez, A. (2010). *Secure three-party key distribution protocol for fast network access in EAP-based wireless networks*, *Computer Networks* 54: 2651 – 2673.

N. Nasser, A. Hasswa & H. Hassanein (2006). *Handoffs in Fourth Generation Heterogenous Networks*, *IEEE Communications Magazine* vol. 44(10): pp. 96–103.

P. Calhoun, G. Zorn, D. Spence & D. Mitton (2005). *Diameter Network Access Server Application*. IETF RFC 4005.

P. Calhoun & J. Loughney (2003). *Diameter Base Protocol*. IETF RFC 3588.

P. Eronen, T. Hiller & G. Zorn (2005). *Diameter Extensible Authentication Protocol (EAP) Application*. IETF RFC 4072.

R. Dantu, G. Clothier & Anuj Atri (2007). *EAP methods for wireless networks*, *Elsevier Computer Standards & Interfaces* vol. 29: pp. 289–301.

R. Housley & B. Aboba (2007). *Guidance for Authentication, Authorization, and Accounting (AAA) Key Management*. IETF RFC 4962.

R. M. Lopez, A. Dutta, Y. Ohba, H. Schulzrinne & A. F. Gomez Skarmeta (2007). *Network-Layer Assisted Mechanism to Optimize Authentication Delay during Handoff in 802.11 Networks*, *Proc. of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, ACM Mobiquitous 2007*, ACM, Philadelphia, USA.

R. Marin, J. Bournelle, M. Maknavicius-Laurent, J.M. Combes & A. Gomez-Skarmeta (2006). *Improved EAP keying framework for a secure mobility access service*, *Proc. of International Wireless Communications & Mobile Computing Conference 2006, IWCMC 2006*, Vancouver, British Columbia, Canada, pp. 183–188.

S. Pack & Y. Choi (2002). *Fast Inter-AP Handoff using Predictive-Authentication Scheme in a Public Wireless LAN*, *Proc. of IEEE Networks 2002 (Joint ICN 2002 and ICWLHN 2002)*.

S. Winter, M. McCauley, S. Venaas & K. Wierenga (2010). *TLS encryption for RADIUS*. IETF Internet-Draft.

T. Aura & M. Roe (2005). *Reducing Reauthentication Delay in Wireless Networks*, *Proc. of 1st IEEE Security and Privacy for Emerging Areas in Communication Networks, SECURECOMM 2005*, IEEE, Athens, Greece, pp. 139–148.

T. Clancy, M. Nakhjiri, V. Narayanan & L. Dondeti (2008). *Handover Key Management and Re-authentication Problem Statement*. IETF RFC 5169.

T. Dierks & C. Allen (1999). *The TLS Protocol Version 1.0*. IETF RFC 2246.

Taniuchi, K., Ohba, Y., Fajardo, V., Das, S., Yuu-Heng, M. T. C., Dutta, A., Baker, D., Yajnik, M. & Famolari, D. (2009). *IEEE 802.21: Media independent handover: Features, applicability, and realization*, *IEEE Communications Magazine* 47(1): 112 –120.

V. Narayanan & L. Dondeti (2008). *EAP Extensions for EAP Re-authentication Protocol (ERP)*. IETF RFC 5296.

Y. Ohba and A. Yegin (2010). *Pre-Authentication Support for the Protocol for Carrying Authentication for Network Access (PANA)*. IETF RFC 5873.

Y. Ohba, Q. Wu & G. Zorn (2010). *Extensible Authentication Protocol (EAP) Early Authentication Problem Statement*. IETF RFC 5836.

Z. Cao, H. Deng, Y. Wang, Q. Wu & G. Zorn (2011). *EAP Re-authentication Protocol Extensions for Authenticated Anticipatory Keying (ERP/AAK)*. IETF Internet Draft, raft-ietf-hokey-erp-aak-06.

2

IP and 3G Bandwidth Management Strategies Applied to Capacity Planning

Paulo H. P. de Carvalho, Márcio A. de Deus and Priscila S. Barreto
Departament of Electrical Engineering, Departament of Computer Science
University of Brasilia
Brazil

1. Introduction

This chapter discusses the application of methodologies to plan and design IP Backbones and 3G access networks for today's Internet world. The recent trend of the multi-frequency band operations for mobile communication systems requires increasingly bandwidth capacity in terms of core and access. The network planning task needs mathematical models to forecast network capacity that match the service demands. As the nature of network usage changed, to explain and forecast the network growth, new methods are needed. In this chapter, we will discuss some strategies to optimize the bandwidth management of a real service provider IP/MPLS backbone and later we will propose a method for traffic engineering in a national IP backbone.

Currently, all telecommunications networks are using IP packets to transport several kind of services. The industry has called this integration as IMS (IP Multimedia Subsystem) in 3G technologies. One important challenge is how to implement this desirable integration with the lack of well known mathematical models to perform capacity planning and forecast the network needs in terms of growth and applications demands. In other way, the main question is how to deliver the required level of service for all kind of applications using the same structure but with different types of traffic and QoS (Quality of Service) requirements.

Due to the fact that many different services will use the same transport infrastructure, the Quality of Service can also be described as a result of traffic characterization because the traffic nature per service or at least per application shall be known. As demonstrated in some research papers (Leland et al., 1994; Carvalho et al., 2009), the Erlang model is not able to accurately describe the behavior of Ethernet and Internet traffic. Without the right model, scientific prediction becomes very difficult and therefore, the planning and forecasting tasks become almost impossible. The above research works verified that the Poisson traffic model is not able to explain the IP traffic dynamics and this implies that the capacity planning tasks for integrated services will need new methodologies. Some models have been used with superior performance to achieve these goals, the self-similar or mono-fractal model show acceptable results in several situations (Carvalho et al., 2007).

Several works show that the multifractal models are particularly promising for multimedia networks (Riedi et al., 2000; Abry, 2002; Fonseca, 2005; Deus, 2007). The traffic

engineering task is valuable to optimize the network resources such as links, routing and processing capacity. One important issue in the traffic engineering task is that the capacity planning forecasting may be for medium long periods (or more than one year), due the fact is not easy to increase long distance link capacities in small periods of time. This problem is much more valuable when the coverage area income is not proportional to the area, as in countries like Brazil, China, Russia in which large areas not necessarily economically attractive.

2. Network planning

The planning task is fundamental to optimize resource utilization. The Fig. 1 describes, from an industry point of view, a complete feasible telecommunications planning cycle. The inputs are the service demands, described as all type of products/services needs per region and also per customer. The physical and logical inventory are very important to be accurate in terms of transmission mediums such as fiber or radio, demographic dispersion, network elements complete description, management assets, and other important physical and logical information.

In terms of innovation, the approach is to use new technologies to achieve new degrees of service delivery; this function shall be used as a complement for planning and forecasting purposes. Other very important function is the economic variables to calculate the return of the investments (ROI) and all other related costs (fixed and variable). All information about traffic usage will be collected and sampled depending on the nature of the service and will have a fast track for immediate operations and decision-making, normally every 5 minutes. For long-term planning these samples will be aggregate in hours, days and weeks.

The functions in Figure 1, in terms of long term capacity will be used to achieve the capacity to deliver new services allowing network expansion related to the inputs, generating new routing and topology and other capacity needs, as described in Figure 1. The traffic engineering function is used in real-time, under human supervision, sometimes even when some modification in terms of routing is proposed by an algorithm. Sometimes, this could not be feasible in practice because network stability is more important in operational environments (Carvalho et al., 2009; Evans & Filsfils, 2007).

The peering agreements will be done as a function of the outputs and also observing the commercial issues. In this way, many service providers have a peering committee to approve new peering interconnections, which has not only a technical importance as well as a marketing approach. The capacity outputs will generate purchasing activities; this will be done by an engineering implementation function. The main objective is to have an operational network, providing all kind of facilities and desirable services.

Along with the massive growth of the Internet and other applications, an increasing demand for different kinds of services for packet switching networks is important. Nowadays, these networks are expected to deliver audio and video transmissions with quality as good as that of a circuit switching network. In order to make it possible, the network must offer high quality services when it comes to bandwidth provisioning, delay, jitter and packet loss.

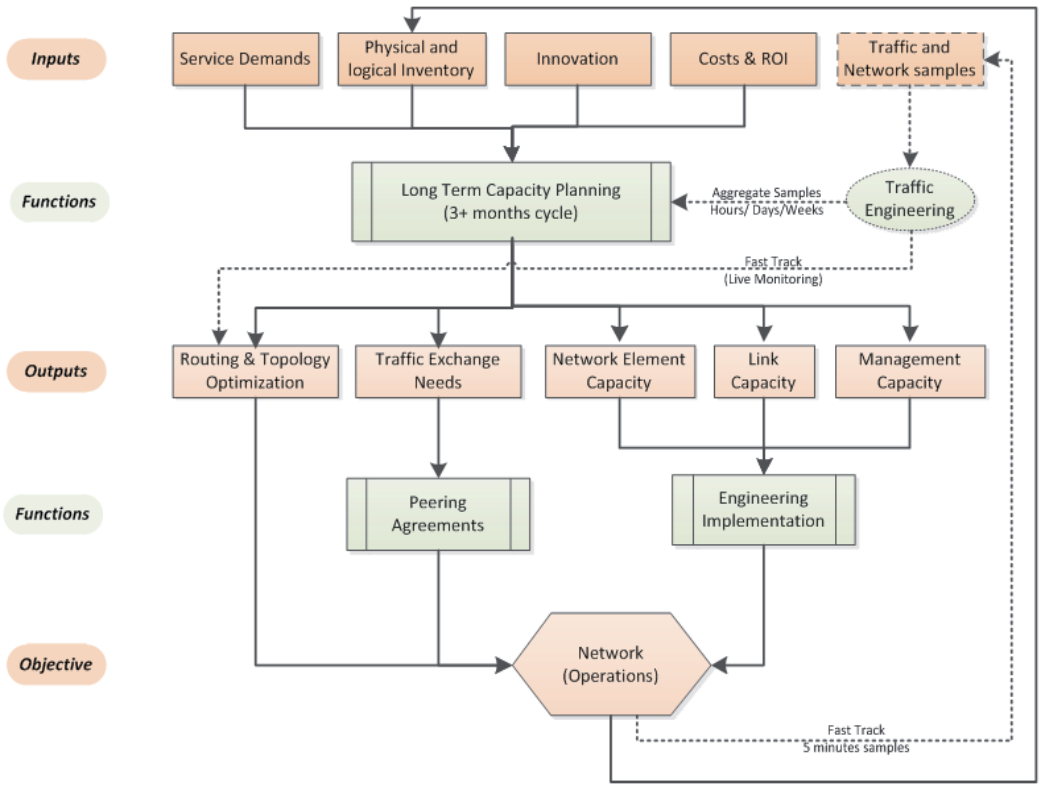


Fig. 1. Telecommunications Industry Planning Process. Adapted from (De Deus, 2007; Evans & Filsfils, 2007).

The processes of traffic characterization and modelling are very important points of a good network project. A precise traffic modelling may allow the understanding of a physical network problem as a mathematical problem whose solution may be simpler. For example, the use of traffic theory suggests that mathematical models can explain, at least for some confidence degrees, the relationship between traffic performance and network capacity (De Deus, 2007; Fonseca, 2005).

The next sections will provide an example on a 3G network using traffic samples to study the planning and project deployment phases. The network described in our study runs with more than 1 million attached 3G costumers with national coverage. In this network, we collected traffic in July 2009 in three different locations (Leblon, Barra da Tijuca and Centro) in Rio de Janeiro. In this way, the first step was to classify the traffic per application. The second step was to characterize the traffic using a procedure based on self-similarity (Clegg, 2005) or multifractal analysis (Carvalho et al., 2009). These results were used as basis for proposing a method to manage the traffic in the network.

To manage the traffic demands, we deployed a traffic engineering concept that divides the traffic across the network through tunnels. The bandwidth was monitored and in the observed period, we collected metrics that were used as inputs to decide how to configure new parameters that may fit the incoming needs. An ILEC (incumbent local exchange

carrier) service provider of IP traffic was used to collect real network traces and we simulated a similar architecture of this network using the OPNET Modeler tool.

A 3G with a Metro Ethernet access was also analysed. The analysis considered a per application separation of traffic. The statistical analysis was done using a self-similarity approach, calculating the Hurst parameter using different calculation methodologies (Abry et al., 2002). Some multifractal analysis was also done as a tool to better choose the time scale.

The results show that the proposed method is able to generate better results in terms of an on-line traffic engineering control and also to provide key information to long term capacity planning cycles. The Traffic Engineering function is detailed using some network simulations examples. Finally, some long term forecasting and short term traffic engineering proposal was done in a 3G networks.

2.1 Traffic modelling in multimedia networks

The traffic modelling and its application to real traffic in operational networks, allows the implementation of research platforms that simulate future or real network critical conditions, which is particularly interesting for huge service providers. Injecting traffic series generated accordingly to mathematical models may help to evaluate several conditions in a network and certainly this may help to develop more accurate capacity planning models regarding specific QoS requirements. Such procedures also facilitate the creation of management strategies. A large number of tools on the Internet provide traffic analysis, like TG (TG), NetSpec (NetSpec), Netperf (Netperf), MGEN (MGEN) and D-ITG (D-ITG) and GTAR, Gerador de Tráfego e Analisador de QoS na Rede (Carvalho et al., 2006), FracLab (FracLab, 2011).

To model the traffic in integrated networks is necessary the use of mathematical models that allow, from its base, to infer the impact of traffic on network performance. The efficient characterization of traffic will be given by the degree of accuracy of the model in comparison with the real traffic statistical properties.

In our work, the characterization of the traffic is used as a key element in the design of complex telecommunications systems. Once characterized, the traffic on different time scales can be used in network simulations. The simulation process can reproduce the behaviour of traffic by application type, for parts of the network, by customer group or interconnections with other networks, opening the possibility to increase the knowledge of the network and making possible a better control of resources.

2.2 Poisson and erlang model

The use of the Internet to transmit real-time audio and video flows increases every day. Some of these applications are transmitted at a constant rate. This kind of traffic results by sending one packet every 1/Tx seconds, where Tx is the rate of transmission in packets per second, defined by the type of the application.

In circuit switched networks, a very successfully model is based on the Poisson distribution. The Poisson traffic is characterized by exponentially distributed random variables to

represent the inter-packet times. The Erlang model, broadly used in telephony systems has been successfully used for capacity planning for many years and is based in the premise that a Poisson distribution describes the traffic in this type of network.

The Poisson model was considered accurate in the early years of the packet switched networks and was heavily used for capacity planning. In the early 90's, the work of Leland(Leland et al., 1994) proved that the behavior of the Ethernet traffic was considerably different than Poisson traffics mainly regarding self-similar aspects with long-range dependence, which is not well described by short memory processes. In practice, the packet switched networks that were planned using the Poisson model, normally had an overprovision in links capacity to comply with the lack of accuracy of the model. Considering the different works about capacity planning following the work of Leland, the heavy-tail models were considered more accurate to describe the traffic in packet switched networks and consequently, they appeared as a better choice.

2.3 Self-similar

One kind of traffic that appears often in wideband networks is the burst traffic. It can be generated by many applications such as compressed video services and file transfers. This traffic is characterized by periods with activity (on periods) and periods without activity (off periods). Moreover, as proved in (Perlingeiro & Ling, 2005), (Barreto, 2007), it is possible to generate self-similar traffic by the aggregation of many sources of burst traffics that presents a heavy-tailed distribution for the on period.

The self-similar model defines that a trace of traffic collected at a time scale has the same statistical characteristics that an appropriately scaled version of the traffic to a different time scale (Nichols et al., 1998). From the mathematical point of view, the self-similarity of a stochastic process in continuous time is defined as shown in Equation 1, which defines a process in continuous time X (t) as exactly self-similar.

$$X(t) \overset{d}{=} a^{-H} X(at), a > 0 \tag{1}$$

The sample functions of a process X(t) and its scaled version of the $a^{-H}X(at)$ obtained by compressing the time axis by the factor amplitudes "a" , can not be distinguished statistically. Therefore, the moments of order n of X(t) are equal to the moments of order n of X (at), scaled by a-Hn. The Hurst parameter, H is then a key element to be identified in the traffic. For self-similar traffic, the H is greater than 0.5 and less than 1. For a Poisson traffic this value is close to 0.5. Experimental results show that this same parameter in operational networks (Perlingeiro & Ling, 2005; Carvalho et. Al., 2007) has values between 0.5 and 0.95. Then, the parameter H may be a descriptor of the degree of dependence on long traffic (Zhang et al.; 1997).

The aforementioned Hurst parameter plays a major role on the measurement of the self-similarity degree. The closer it is of the unity, the greatest the self-similarity degree. One of the most popular self-similar processes is the fractional Brownian motion (fBm), which is the only self-similar Gaussian process with stationary increments. The increments process of the fBm is the fractional Gaussian noise (fGn). To generate the traffic, we first create a fGn

sequence based on the method presented in (Norros, 1995). Each sample of the sequence represents the number of packets to be sent on a time interval of size T. The size of the time interval and the mean of the sequence generated will depend on the traffic rate.

2.4 Multifractal traffic

As self-similar models, multifractals are multiscale process with rescaling properties, but with the main difference of being built on **multiplicative** schemes(Incite, 2011). In this way, they are highly non-Gaussian and are ruled by different limiting laws than the additive CLT (Central Limit Theorem). Therefore, multifractals can provide mathematical models to many world situations such as Internet traffic loads, web file requests, geo-physical data, images and many others. The Hölder function is defined by the h(t) function.

In the self similar model, also called as monofractal, the Hurst parameter is a global property that quantifies the process changes according to changes in the scale. For multifractal traffic, however, the Hurst parameter becomes less efficient in this characterization and another metric is needed to perform the scaling analysis of the sample regularity.

There are several ways to infer the scaling behavior of traffic, one way is widely used by local singularities of the function. A singular point is defined as a point in an equation, curve, surface, etc., which have transitions or becomes degenerate (Ried et al., 2000). It is quite common that the singular points of the signal containing essential information on network traffic packets.

In order to identify the singularities of a signal, it is necessary to measure the regularity of the same point, which will reflect in burst periods occurring at all traffic scales. In (Gilbert & Seuret, 2000) some examples can be found about the point and the exponents of the local Hölder values making possible to check the degree of uniqueness of network traffic.

According to Veira, (Veira et al., 2000) the Hölder exponent is capable to describe the degree of a singularity. Considering a function $f : R \rightarrow R$, with x_0 as real number, and α a stricted real positive number. It can be assumed that f belongs to $C_\alpha(x_0)$ if a polynomial P_m with degree $n < \alpha$, as shown in (2).

$$\left| f(x) - Pm(x - x_0) \right| \leq C \left| x - x_0 \right|^\alpha \tag{2}$$

As described in (Ludlam, 2004) a multifractal measure P can be characterized by calculating the distribution f (α) , known as the multifractal, or singularity, spectrum where α is the local Hölder exponente (Clegg, 2005 ; Castro e Silva, 2004 ; Vieira, 2006). This measure can be also shown as a probability density function P (x), in this case, the local Hölder exponente (; Gilbert & Seuret, 2000) is defined ad in (7).

$$\alpha (x) = \lim l \rightarrow 0 \log P (\mathcal{B} (l , x)) \log l \tag{3}$$

where $\mathcal{B} (l , x)$ is a box centred at x with radius l , and P ($\mathcal{B}$) is the probability density integrated over the box $\mathcal{B}$. It describes the scaling of the probability within a box, centred on a point x , with the linear size of the box.

Each point x of the support of the measure will produce a different α (x) , and the distribution of these exponents is what the singularity spectrum f (α) measures. The points for which the Hölder exponents are equal to some value α form a set, which is in turn a fractal object. The fractal dimension of this set can be calculated, and is a function of α , namely f (α).

As described in (2), a function $f(x)$ satisfies the Hölder condition in a neighborhood of a point, where c and n are constants, as in (4).

$$x_0 \text{ if } |f(x) - f(x_0)| \leq c\,|(x-x_0)|^n \tag{4}$$

And a function $f(x)$ satisfies a Hölder condition in an interval or in a region of the plane, for all x and y in the interval or region, where c and n are constants, as in (5).

$$|f(x) - f(y)| \leq c\,|x - y|^n \tag{5}$$

3. Traffic characterization

The process of traffic characterization is a preponderant point of a feasible network project. In this section a traffic characterization framework is described. The characterization intends to describe a step by step procedure, which may be useful to understand the behavior of traffic in large networks using a mathematical model as a tool to achieve good planning. One difficult issue to characterize traffic in IP networks is the changing environment due to new applications and new services that are appearing constantly. This implies that the

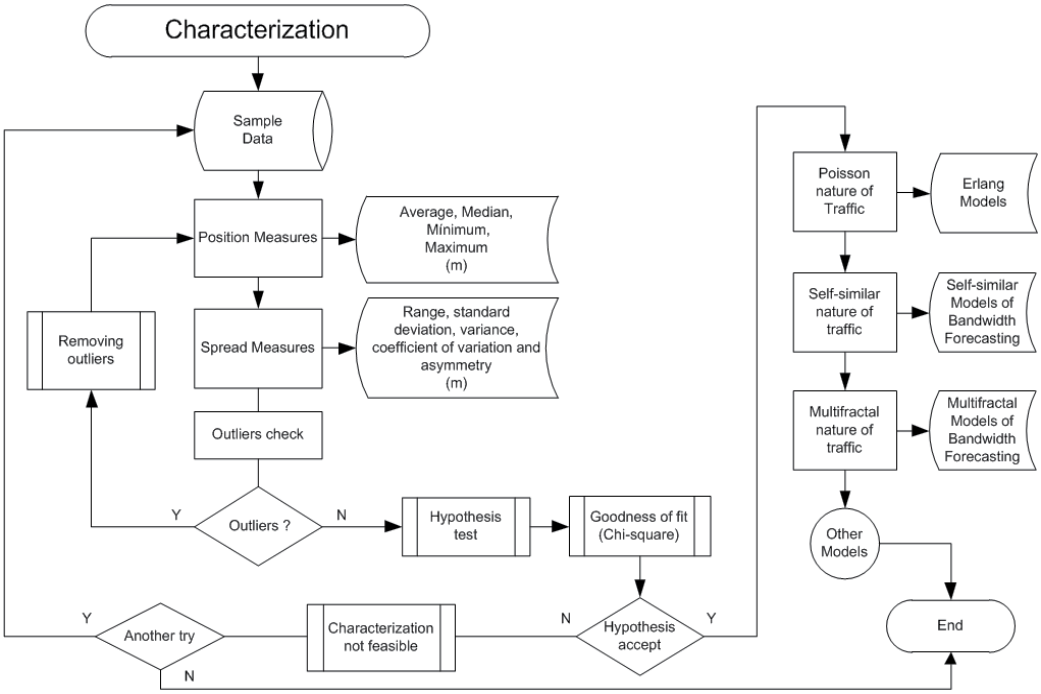


Fig. 2. Characterization process.

characterization used in real environments shall considerer the evolution and the amount of variation in the types of services, including not well known agents as social behavior and emerging applications.

The efficiency in traffic characterization is given by the model accuracy when compared with real traffic measures. As said by (Takine et al., 2004) a traffic model can only exist if there is a procedure for efficient and accurate inference for the parameters of the same mathematical structure. The traffic characterization is the main information source for the correct mathematical interpretation of network traffic. Once characterized, the traffic may be reproduced in different scales and periods and inserted into network simulators.

Figure 2 shows a complete characterization flow to optimize planning. This procedure was implemented in the GTAR (Barreto, 2007) simulator, developed within our research.

4. Experimental analysis

4.1 Analysis of an IP network

The first network to be evaluated is a Brazilian Service Provider in Brazil, with more than ten million PSTN (Public Switched Telephone Network) subscribers and more than one million ADSL as well. The IP network is shown Figure 3 each access layer is a PPPoX router capable called BRAS(Broadband Router Access Server).

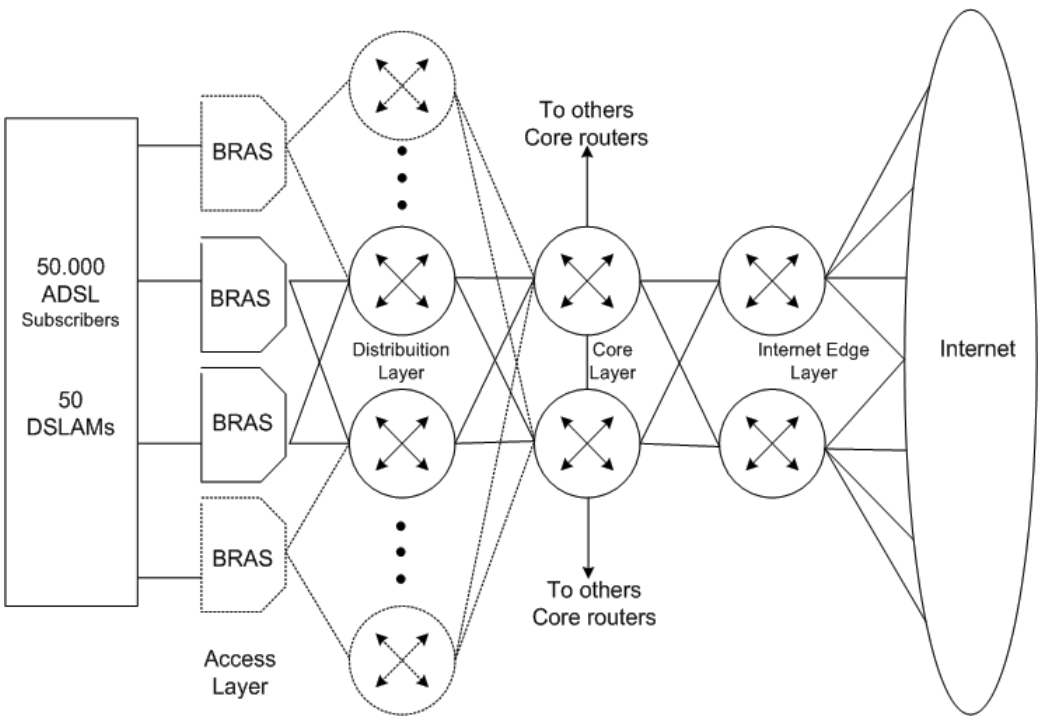


Fig. 3. Testbed Network Architecture with 40% of simultaneous attached subscribers at least, all IP/MPLS interface 1 or 10 Gigabit Ethernet, also for long distance. (De Deus, 2007).

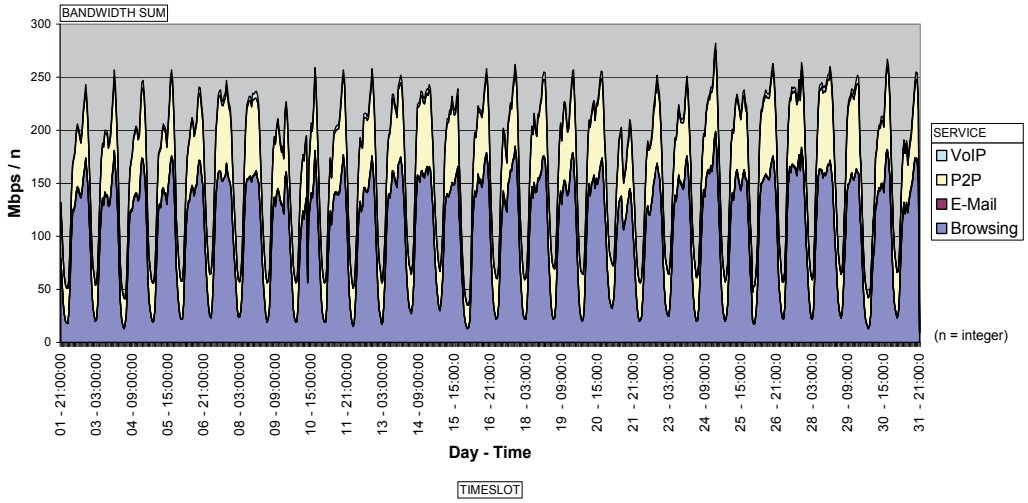


Fig. 4. Downstream traffic "on peak" and "off peak". The rate is normalized, 31 days sampled (De Deus, 2007).

Figure 4 and 5 shows the downstream traffic collection results for a 31 days period. The most important source of traffic is the HTTP(Browsing) following by P2P applications(e-Donkey, Bitorrent, Kazaa). In Figure 10, the same analysis is made for a 24 hours period.

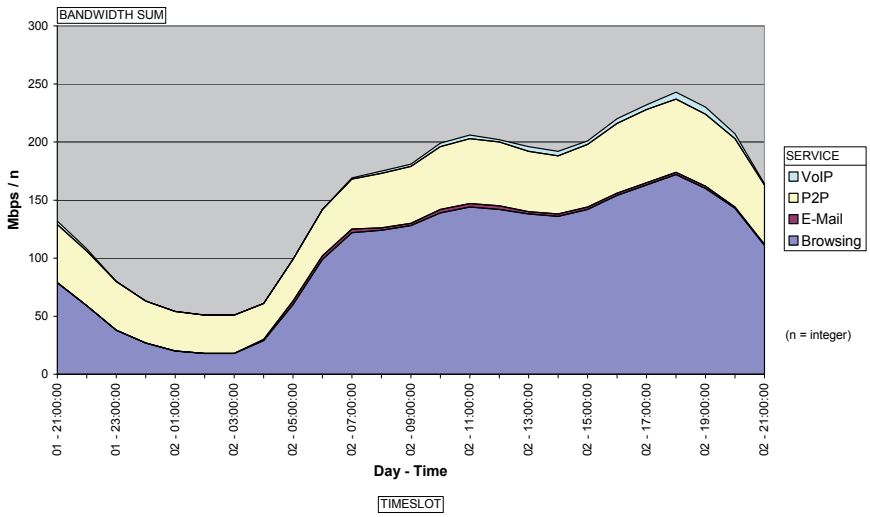


Fig. 5. Downstream traffic "on peak" and "off peak". Traffic rate is normalized, 24 hours sampled (De Deus, 2007).

Figure 6 shows the packet size probability distribution. Less than 100 Bytes packets have 50% of probability. These samples are from a real network with Internet traffic of 4 million xDSL subscribers, demonstrating the very large use of voice packets even when using http flows. This happens mainly because of applications such as SKYPE.

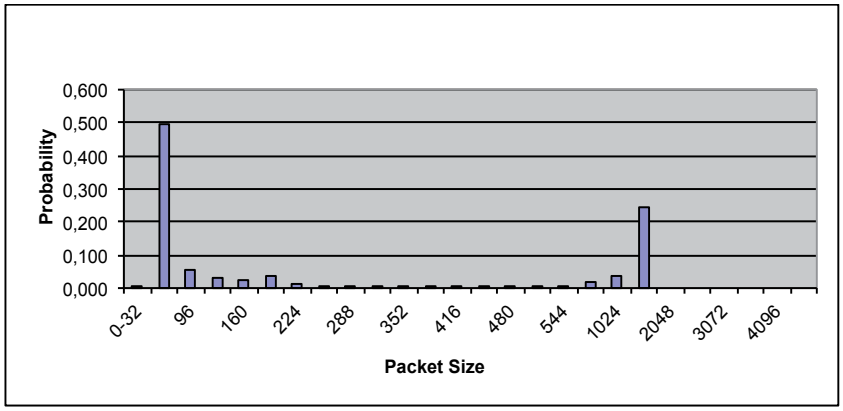


Fig. 6. Packet Size Probability Distribution (De Deus, 2007).

Table 1 shows and per application anaylis of traffic in which the Hurst parameter was calculated with two different methods (De Deus, 2007). For real time traffic, the Hurst parameter calculation demands attention because in some cases if statistical process does not have a representative long range dependence characteristic the parameter may be wrongly interpreted. Another issue is the trend present in the periodic traffic. For a more accurate estimation, the cycle regularity is removed to delete all observed trends.

Day	Hurst (*Variance-Time Plot*)	Hurst (Kettani-Gubner)	Chi-Square (Gaussian Distribution)
1	0,843	0,895	31,042
2	0,812	0,878	71,299
3	0,901	0,926	38,146
4	0,9	0,934	52,569
5	0,815	0,879	32,549
6	0,816	0,904	62,042
7	0,865	0,906	17,91
8	0,87	0,916	39,653
9	0,907	0,935	28,028
10	0,867	0,919	21,785
11	0,869	0,906	27,167
12	0,671	0,861	35,778
13	0,878	0,909	36,208
14	0,839	0,894	44,604
15	0,874	0,907	30,611
16	0,753	0,85	23,292
17	0,851	0,914	40,299

Day	Hurst (*Variance-Time Plot*)	Hurst (Kettani-Gubner)	Chi-Square (Gaussian Distribution)
1	0,915	0,948	63,549
2	0,942	0,962	49,771
3	0,937	0,963	45,25
4	0,902	0,935	28,243
5	0,902	0,928	20,708
6	0,901	0,942	30,181
7	0,939	0,964	43,258
8	0,932	0,964	52,354
9	0,937	0,968	39,007
10	0,922	0,948	38,576
11	0,86	0,926	37,5
12	0,904	0,942	33,84
13	0,937	0,965	55,799
14	0,922	0,958	46,972
15	0,935	0,963	46,757
16	0,935	0,967	49,986
17	0,933	0,964	37,285

Table 1. HTTP and P2P Hurst parameter estimation for 5 minutes average.

In Table 1 is shown the estimation of the H parameter for the HTTP (Hyper Text Transfer Protocol) applications. As can be seen, the H relies value between 0.67 and 0.93, which also shows a higher degree of self-similarity, considering that the lower value appears just in one day. For the P2P applications, the H parameter relies between 0.86 and 0.96.

The estimation of the the Hurst parameter in Table 1 uses three different methods: the Variance-Time Plot Method, the Kettani-Gubner Method (Clegg, 2005), (Barreto, 2007). Also a Chi-squared analysis was made as a non-parametric test of significance (Perlingeiro, 2006), (De Deus, 2007), (Clegg, 2005) due to the fact that it is necessary to verify the distribution similarity. The statistical significance test allows, with a certain degree of confidence, the acceptance or rejection of a hypothesis, as shown in Figure 7. The sampled links had a load, in the worst case around 70%.

Figure 7 shows the Hölder calculation for the traffic. The conclusion in fact is that the traffic is self-similar and monofractal, when the measurement is done in a 5 minutes per sample.

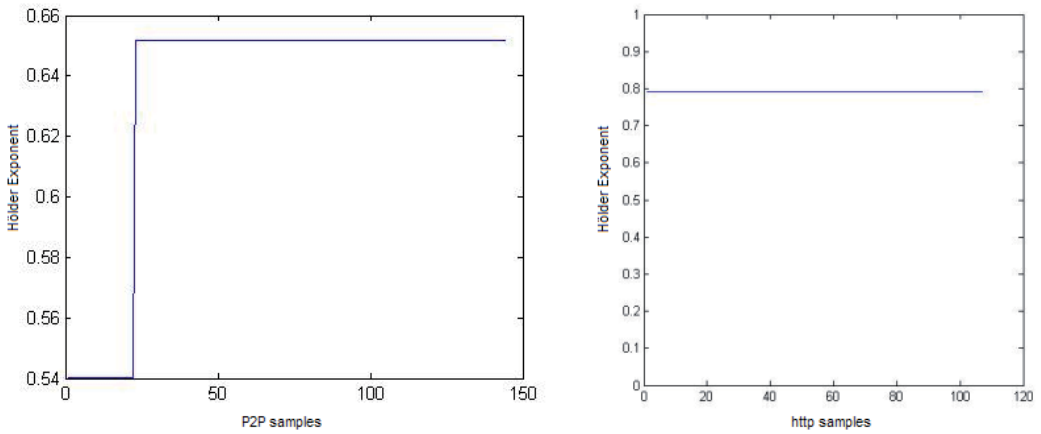


Fig. 7. P2P and http 5 minutes samples, Hölder exponent using the local Hölder Oscillation Based method [fraclab].

4.1.1 Bandwidth control strategies for the IP network

Figure 8 shows the proposal of a real-time network forecast. First, in the network the samples are collected. Then the traffic is classified per application. The estimation and a characterization of the parameters of collected samples are calculated. These parameters are used as input to a traffic forecast tool based on a mathematical traffic model which intends to find the sub-optimal capacity of the link for that traffic load, considering its self-similarity nature.

The objective is to use these parameters as inputs of a simulation tool to forecast the traffic and feedback in real-time the network to provide a new model to capacity plan in the backbone.

Following Figure 8, first the network samples are collected. Thenext step is the execution of classification procedure per application using tools based on protocols (Destination, Source, Port, Payload types). Next phase is to estimate the parameter (e.g. Hurst, Hölder) that will

be used as input to a traffic forecast tool based on valid models (Norros *at al.*, 2000). The next step is to insert the parameter to a tool that will take a decision of how the auto-configuration will be done and a configuration of the element abstracting the vendor (e.g. Juniper, Cisco, Huawei). In figure 7 the example of application of the feedback process is described using the auto configuration tool to change the tunnel characteristics, that will use the proposed framework in Figure 8, as an example of setting up an outstream traffic marked as Diffserv.

If the traffic can be characterized as asymptotical self-similar or monofractal or multifractal some ready prediction models based (e.g. fBm, MWM, MMW) can be used. The core idea is that using only some parameters the mathematical calculus can be feasible at real time, as shown in Figure 14.

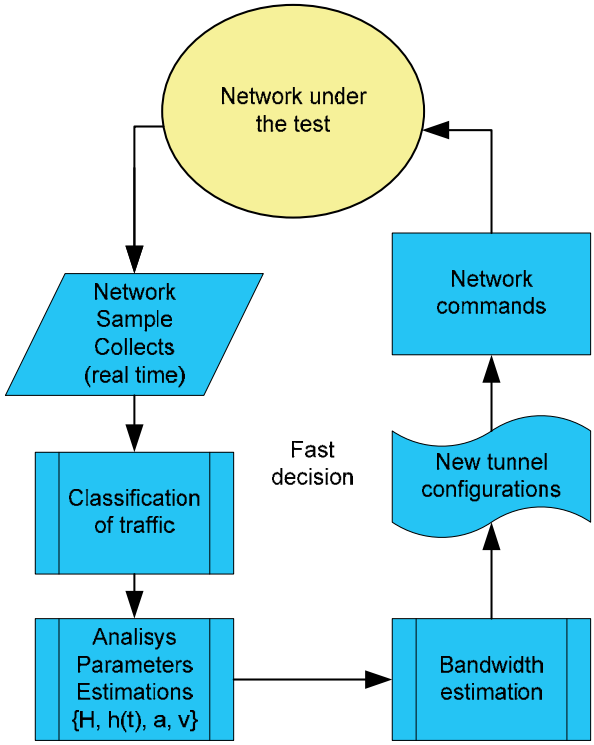


Fig. 8. Proposal of a network real-time forecast framework with bandwidth estimation.

In this case, the tunnels are configured using the self-similarity bandwidth estimators, as described in (Carvalho, 2007). The traffic needs to be marked as the DiffServ and will be injected per tunnel as the auto configuration tunnel selection.

There are several methods used to estimate bandwidth. The method used in our example is the FEP(Fractal Envelope Process). This model has a good performance for long range dependence with a high degree of confidence in the quasi-Real Time estimation (De Deus, 2007).

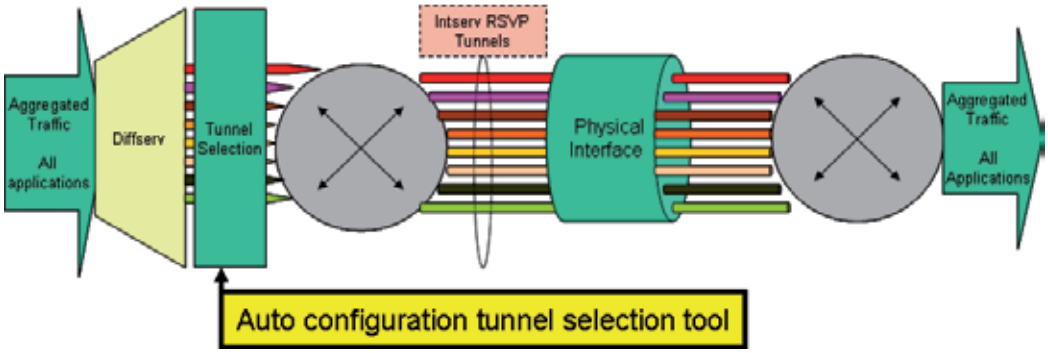


Fig. 9. Tunnel selection between two routers using Diffserv and Inteserv to select the specific tunnel.

The bandwidth estimation most accepted definition, currently known, use a concept introduced by (Kelly et al.,1996), where there is a direct dependency on buffer size and time scales related to the buffer overflow possibility. The concept is shown in (6) where X[0, t] is the amount of bits that arrive in an interval [0, t], considering that X[0, t] has stationary increments. The letter b is the buffer size and t time or scale, BP is the capacity in bits per second.

$$BP(b,t) = \frac{\log E\left[e^{bX[0,t]}\right]}{bt} \quad 0 < b,t < \infty \tag{6}$$

Based on this theory, several bandwidth estimators have been proposed and evaluated for its effectiveness and complexity of evaluation. In (Fonseca et al., 2005) an evaluation of the FEP estimator model (Fractal Envelope Process) was developed with good results, for use in high speed networks.

Equation (7) represents the FEP process estimation where the K is the buffer, a is the average, H is the Hurst parameter, σ is the standart devitation and Ploss represents the probability of packet loss when a buffer overflow. This is only valid when 0.5 < H < 1.

$$EN = \overline{a} + K^{\frac{H-1}{H}} * \left(\sqrt{-2 * \ln(P_{loss})} * \sigma\right)^{\frac{1}{H}} * H(1-H)^{\frac{1-H}{H}} \tag{7}$$

Using (7) and correcting with (8) and (9), some curves are plotted in different time scales in Figure 10 (FEP Estimator and FEP Model). The best results are with 5 and 1 minutes, achieving the most next to average but still providing a good service with no delay, jitter or packet loss. The "Modelo FEP" f_{op} means the dynamic calculation per hour, the "Tunel P2P constante" and "Tunel HTTP constant"means the estimation with a Poisson Distribution Estimator, the P2P and HTTP means the dynamic bandwidth calculation.

$$f_{op} = \frac{2}{5}\frac{EN}{\sqrt{b'L}} \; if \; 0.5 < H \leq 0.7 \tag{8}$$

$$f_{op} = \frac{2}{75}\frac{EN}{\sqrt{b'L}} \; if \; 0.7 < H < 1 \tag{9}$$

The f_{op} is calculated based on (Perlingeiro & Ling, 2005) study as shown in (8) and (9), where EN is from (7) and b' is the normalized buffer ($b'=b/b_0$), where b is the buffer and b_0 minimum possible buffer size, L is the burst factor.

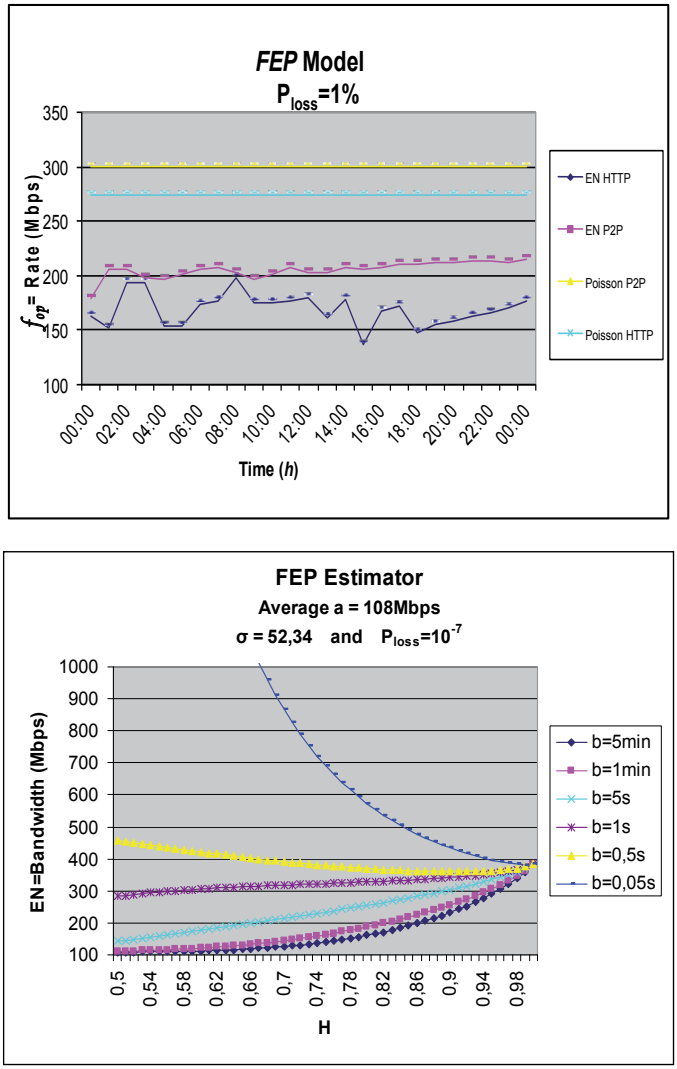


Fig. 10. Bandwidth estimation curves using the FEP method.

The FEP Model shown in Figure 10 uses a dynamic tunnel configurator as shown in Figure 9, denoting a better usage of the total available bandwidth. In the figures it appears that when a constant calculated bandwidth is used, more bandwidth is required. In the same way, the FEP Estimator shows that as much aggregated the traffic will be in any time scale,

the difference will be minimum. In the other hand, when going to small time scales .05, .5 or 1 seconds, there is a trend in super estimation, proportional to the diminishing of the Hurst parameter.

As shown in many works (Leland et al., 1994), (Abry et al., 2002), (Carvalho et al., 2009), the Hurst parameter can show an accurate and single way to determinate the self-similarity. The Erlang model is very useful because its simplicity. A traffic engineer only needs to have some little information about service demand such as Retention time, blocking Probability, Number of Calls in the maximum usage hour to have the traffic and number of channels or resource needed.

The curves in Figure 10 show the possibility to have something, not so easy as Erlang model, but also possible to be achieved as a traffic model when a self-similar characterization is feasible. Also, the multi fractal model can also help to understand this same traffic in smaller scales, or in some case depending the traffic nature.

4.2 Analysis of a 3G network

The second evaluated network is a brazilian 3G network. This network runs with more than 1 million attached 3G costumers with national coverage. The traffic samples were collected in July, 2009 in three different locations (Leblon, Barra da Tijuca and Centro) in Rio de Janeiro. Two monitors were located in the network to collect the traffic, as shown in Figure 11.

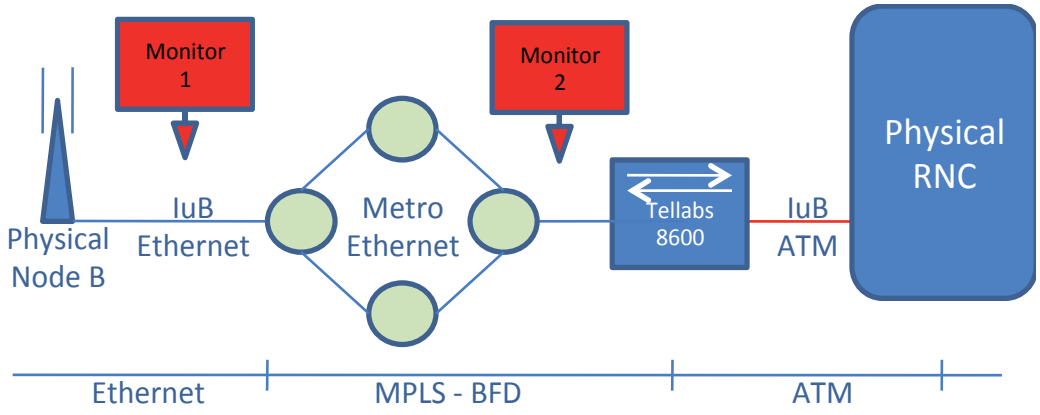


Fig. 11. 3G Network.

The main objective in this section is to investigate planning and project deployment phases based on traffic characterization. The first step is to classify the traffic per application. The second is to characterize the traffic using a procedure based on self-similarity (Clegg, 2005) or multifractal analysis (Carvalho *et al*, 2009).

Figure 12 shows the network topology for the Ethernet physical node B (ATM node) with an ATM-IP router which is responsible to convert ATM to Ethernet(IP). The same situation is found in RNC side where a Tellabs ATM-IP router aggregates all node B physical uplinks,

every one carried through a Metro Ethernet network, with more than 50km radius Rio de Janeiro metropolitan area coverage.

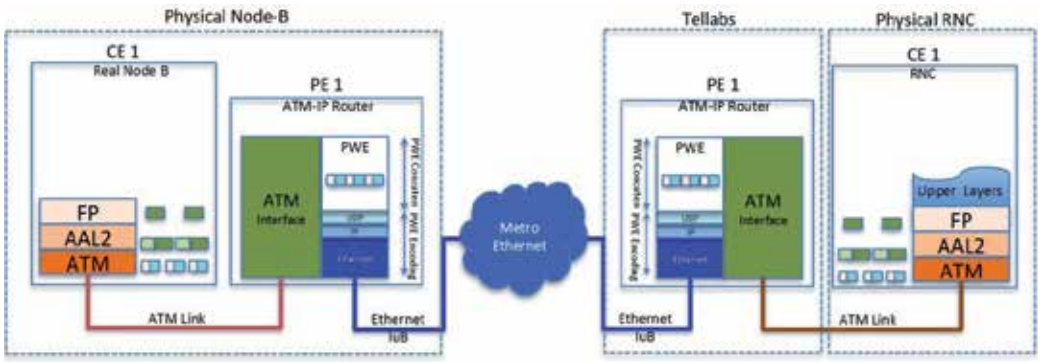


Fig. 12. 3G topology from Node B to RNC.

The first performance analysis of this network found some drawbacks in terms of latency and packet loss and jitter. In Figure 13 (before) is shown the first measures. One detected problem was the high level of broadcasting (ARP included) for this metro Ethernet network, in some periods, more than 80% of all IP traffic.

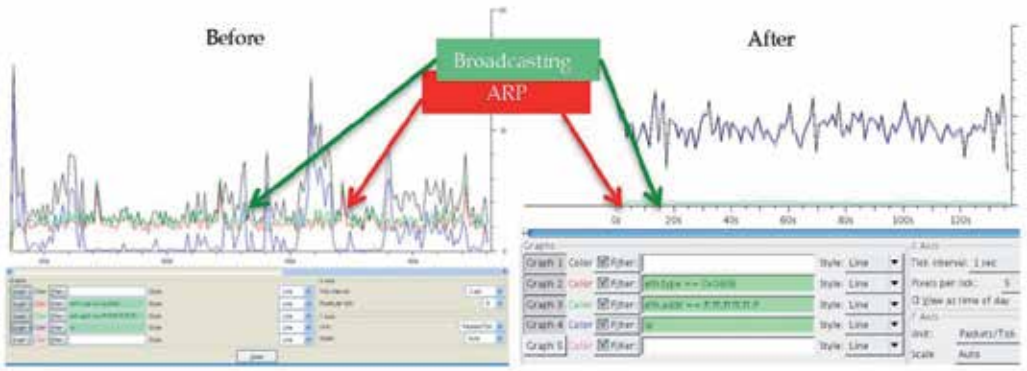


Fig. 13. 3G Traffic analysis (before and after).

As shown in Figure 12, the transport from Node-B to the RNC is performed by a MetroEthernet network that uses also a BFD protocol to track the availability of a Multiprotocol Label Switching (MPLS) Label Switched Path (LSP). In particular, BFD (Aggarwal et al., 2010) can be used to detect a data plane failure in the forwarding path of an MPLS LSP.

LSP Ping is an existing mechanism for detecting MPLS LSP data plane failures and for verifying the MPLS LSP data plane against the control plane, making possible the PseudoWire connections through a MPLS environment.

The problem, in this case, was an architectural design mistake because all Node B uplinks were configured in Level 2 VLANs (OSI Model), with more then 250+ 3G nodes B in the

same IP subnet. The solution for this architectural problem was divide the Node Bs in 20 per subnet, as shown in Figure 8 (after).

This division resulted in diminishing the broadcasting to less than 5%. This problem is very simple in a typical Ethernet topology, but not so easy to be detected when inserted in a 3G network. Ethernet is a protocol designed for local area purposes; the MEF (Metro Ethernet Forum) inserted some signalling standards as a way to simplify the application in metro and long-range use.

Figure 14 shows the traffic trace collected in the 3G network and Figure 15 and Figure 16 show the singularity spectrum and the Hölder function for the 3G samples, showing the possibility to use the multifractal model also to forecast purposes.

This information is important do show this traffic can be characterized as multifractal in small scales of time, but in other hands the bandwidth model for this type of traffic model is also hard to build, because the nature of the traffic. Other important thing to understand is how to insert modifications with make the system not stable. In small scales, huge systems will need a lot of information to compute the bandwidth between to distance nodes.

The use of one model type can be very carefully choose because this could make the Operations Staff make wrong decisions that could result in many downtime.

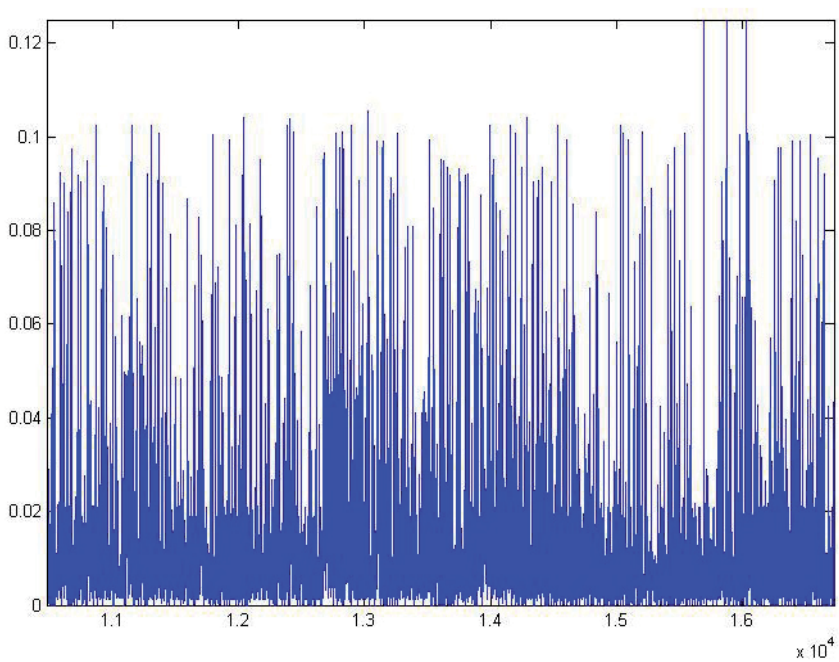


Fig. 14. Normalized 3G Traffic samples (milliseconds time scale).

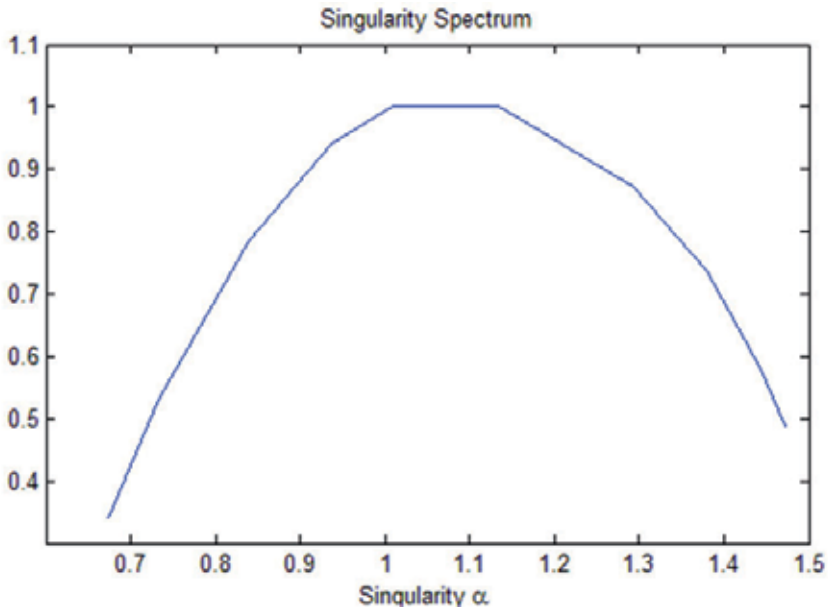


Fig. 15. 3G Traffic multifractal analysis – Singularity Spectrum (milliseconds time scale).

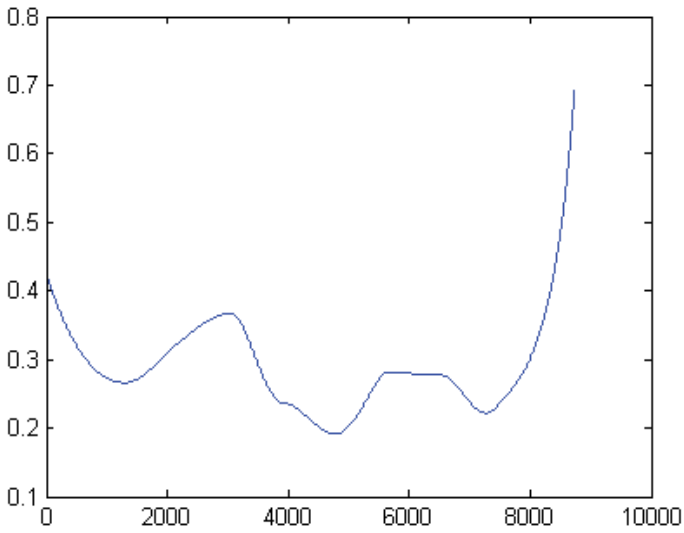


Fig. 16. 3G Traffic multifractal analysis – Hölder function. (milliseconds time scale).

5. Conclusion

This chapter presented an approach and a set of frameworks to characterize traffic and optimize network planning in IP and 3G networks. Based on real traffic measurements, we

characterized the traffic and showed examples of how to apply the proposed frameworks. An special interest of our work has a focus in real operating networks and the examples show the application of the proposed frameworks in these environments.

The traffic characterization procedures for mutilmedia traffic were explained. We provided analyses by collecting different types of traffic and measuring its self-similar or multifractal degrees. All of this work was done with some self-developed (Carvalho et al., 2006) tools and also with some other tools (FRACLAB, 2011; OPNET, 2011).

The traffic models give us a good idea of the traffic behavior. In fact, the models can be valuable tools to the conception, management and sizing of a telecom network, resulting on efficient use of its resources. The operator can plan the growth of the network just to fit the business model, guaranteeing at different moments the efficient use of network resources, guaranteeing, on the other hand, the users satisfaction. In this context, the traffic models can also be used to define alternative policies that, for example, promote the network adaptation in periods with different levels of congestion.

Some very important to considering is how to improve the planning function with a better forecasting (Zukerman et al., 2003)., in terms of long time period for new assets plan and also to implement new products.

Something also very important is how to manage the network resources to have the best optimization possible, this will provide costumer better experience when using and buying exactly their needs.

6. References

Abry P., Baraniuk R.; Flandrin P.,Ried; R., Veitch D. (2002). The Multiscale Nature of Network Traffic Discovery, Analysis and Modeling. IEEE Signal Processing Magazine, 19(3):28-46.

Aggarwal, R; Kompella, K.; Nadeau, T.; Swallow, G. (2010). Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs). RFC 5884. ISSN: 2070-1721, June 2004. IETF Documents.

Avallone, S.; Pescapè, A.; Romano, S.P., Esposito, M.; Ventre, G. (2002). "Mtools: a one-way-delay and round-trip-time meter" 6th WSEAS International Conference, Crete, July 2002.

Barreto, P. S. (2007). Otimização de Roteamento Adaptativo em Redes Convergentes com tráfego autosimilar. Orientador:Carvalho, P. H. P. Tese de Doutorado, UnB.

Carvalho, P. H. P.; Barreto, P. S.; De Deus, M.; Queiroz, B.; Carneiro, B. (2007). A per Application Traffic Analysis in a Real IP/MPLS Service Provider Network. The 2nd IEEE IFIP/ International Workshop on Bradband Convergence Networks(IM2007/BCN2007), IEEE Communications Society , Munich, Germany, 21 a 25 de maio de 2007 ISBN: 1-4244-1297-8. Digital Object Identifier: 10.1109/BCN.2007.372751.

Carvalho, P.H.P.; Barreto, P. S.; Queiroz, B.; Carneiro, B.N. (2006). Modelagem, Geração e Análise de Tráfego em Redes Multiserviços, GTAR, LEMOM, UnB.

Carvalho, P.H.P.; Deus, M. A.; Barreto, P. S. (2009). Effective Bandwidth Allocation for IP/MPLS networks with Multimedia Traffic. In Portuguese : Alocação de Banda Efetiva para Tráfego Multimídia em Redes IP/MPLS. In: I2TS 2009, 2009, FLORIANOPOLIS. 8th International Information and Telecommunication Technologies Symposium, 2009, 2009.

Carvalho, P.H.P.; De Deus, M. A.; Barreto, P. S.; Fraga, T. ; Paiva, V. (2008). Identificação de Características Multifractais para Tráfego de Redes. In: XXVI Simpósio Brasileiro de Telecomunicações (SBrT'08), 2008, Rio de Janeiro, RJ. Anais do XXVI Simpósio Brasileiro de Telecomunicações, 2008.

Castro e Silva, J.L. (2004) "ProCon - Prognóstico de Congestionamento de Redes de Computadores usando Wavelets", Tese de Doutorado, Universidade Federal de Pernambuco, 2004.

Clegg, Richard (2005). A Practical Guide to Measuring the Hurst Parameter, Proceedings of 21st UK Performance Engineering Workshop, School of Computing Science.

D-ITG software (Sep 20, 2011) [Online]. Available: http://www.grid.unina.it/software/ITG/index.php

De Deus, M.A (2007). IP/MPLS Bandwidth Management Strategies for Transport of Integrated Services. Estratégias de Gerenciamento de Banda IP/MPLS para os transporte de Serviços Integrados. Orientador : Carvalho, P.H.P ; Co-orientador : Barreto, P. S.; [Distrito Federal] 2007. xvii, 127p., 210 x 297 mm, ENE/FT/UnB, Mestre, Engenharia Elétrica, Comunicação(2007). Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia.

Evans, J.; Filsfils, C. (2007). Deploying IP and MPLS QoS for Multiservice Networks. Morgan Kaufmann, ISBN-13: 978-0-12-370549-5.

Fonseca, N. L. S.; Drummond, A. C.; Devetsikiotis, M. (2005). Uma Avaliação de Estimadores de Banda Passante Baseados em Medições. Instituto de Computação– Universidade Estadual de Campinas. Department of Electrical and Computer Engineering – North Carolina State University Raleigh, USA

Fraclab (2011), A fractal analysis toolbox for signal and image processing. Available from http://fraclab.saclay.inria.fr/

Gilbert, A. e Seuret, S. (2000). Pointwise Hölder exponent estimation in data network traffic, In ITC Specialist Seminar, 2000.

Huang, J. (2000) "Generalizing 4IPP Traffic Model for IEEE 802.16.3", IEEE 802.16.3c-00/58, Meeting #11, Ottawa, Dec. 2000.

Incite (2011) Available: http://www.ece.rice.edu/INCITE/modeling_synopsis.html

Karagiannis, T; Faloutsos, M.; Riedi, R.H. (2002) "Long-Range Dependence: Now You See It, Now You Don't!" Proc. IEEE Global Telecommunications Conf. Global Internet Symposium, 2002.

Kelly, F.P.; Zachary, S.; Ziedins, I.; editors (1996). Notes on Effective Bandwidth, pages 141–168. Oxford University Press.

Kettani, H.; Gubner, J.A. (2002). "A Novel Approach the Estimation of the Hurst Parameter in Self-similar Traffic", Proceedings of IEEE Conference on Local Computer Networks, Tampa, Florida, November 2002.

Law, A.M; Kelton, W. D. (1991). "Simulation Modeling and Analysis", 2nd ed. New York: McGraw-Hill, 1991.

Leland, W. E; Taqq, M. S.; Willinger, W.; Wilson, D.V. (1994). On Self-similar nature of Ethernet traffic. ACM Sigcomm. Computer Communication.

Ledesma, S.; Liu, D. (2000). "A Fast Method for Generating Self-Similar Network Traffic", Proceedings of the 2000 International Conference on Communication Technologies, Beijing, China, p.54-61, Aug. 2000.

Ludlam, J. (2004). Localisation of the Vibrations of Amorphous Materials. PhD, Thesis Dissertation. Trinity College, Cambridge, UK, 2004. Online at: http://jon.recoil.org/thesis/thesisse11.xml

Melo, E. T. L. (2001). "Qualidade de Serviço em Redes IP com DiffServ: Avaliação através de Medições", 2001.

MGEN software (Sep 20, 2011) [Online]. Available: http://mgen.pf.itd.nrl.navy.mil/

Netspec software (Aug 28, 2011) [Online]. Available: http://www.ittc.ku.edu/netspec/

Netperf software (Jun 7, 2011) [Online]. Available: http://www.netperf.org/netperf/NetperfPage.html

Norros, Ilkka. (1995). On the use of factional Brownian motion in the theory of connectionless networks. IEEE Journal of Selected Areas in Communications, 13(6):953-962.

Opnet (2011), http://www.opnet.com.

Paxson, V. (2000) "Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic", Computer Communication Review, vol.27, p.5-18.

Perlingeiro, F. R.; Ling, L. L.. (2005). Estudo de Estimação de Banda Efetiva para Tráfego auto-similar com variância infinita, SBrT'05, 04-08 de setembro de 2005, Campinas, SP

Riedi, R. H. ; Ribeiro, V. J. ; Crouse, M. S. and Baraniu, R. G. (2000). Network Traffic Modeling Using a Multifractal Wavelet Model. Proceedings European Congress of Mathematics, Barcelona 2000. Department of Electrical and Computer Engineering, Rice University, 6100 South Main Street Houston, TX 77005, USA (NSF/DARPA).

Takine, T.; Okazaki, K.; Masuyama, H. (2004). IP Traffic Modeling: Most Relevant Time-Scale and Local Poisson Property. Department of Applied Mathematics and Physics Kyoto University. (ICKS'04) Informatics Research for Development of Knowledge Society Infrastructure

Taqqu, M. S.; Willnger, M.S.W.; Sherman, B. (1997). "Proof of a fundamental result in self-similar traffic modeling". Computer Communication Review, vol. 27, p. 5-23, 1997.

TG software (Aug 8, 2011) [Online]. Available: http://www.postel.org/tg/

Vieira, F.H.; Jorge, C.; e Ling, L. (2005) Predição Adaptativa do Expoente de Hölder para Tráfego Multifractal de Redes, In XXVIII Congresso Nacional de Matemática Aplicada e Computacional, 2005.

Vieira, Flavio H. T. V. (2006). Contribuições ao cálculo de banda e probabilidade de perda para tráfego multifractal. Tese de Doutorado. Unicamp, 2006.

Zhang, H.F.; Shu, Y.T.; Yang, O. (1997). Estimation of Hurst parameter by variance-time plots. Communications, Computers and Signal Processing, 1997. apos;10 Years

PACRIM 1987-1997 - Networking the Pacific Rimapos;. 1997 IEEE Pacific Rim Conference on Volume 2, Issue , 20-22 Aug 1997 Page(s):883 - 886 vol.2

Zukerman, M.; Neame, T. D.; Addie, R. G. (2003). "Internet Traffic Modeling and Future Technology Implications" Proceedings of Infocom, 2003.

3

eTOM-Conformant IMS Assurance Management

M. Bellafkih[1], B. Raouyane[1,2] , D. Ranc[3], M. Errais[1,2] and M. Ramdani[2]
[1]*Institut National des Postes et Télécommunications, Rabat,*
[2]*Faculté des Sciences et Techniques, Mohammedia,*
[3]*IT Sud Paris, Evry,*
[1,2]*Morroco*
[3]*France*

1. Introduction

QoS management(Raouyane B. et al., 2009) mechanisms as defined by 3GPP can be viewed as a network-centric approach to QoS, providing a signalling chain able to automatically configure the network to provision determined QoS to services on demand and in real time, for instance on top of a DiffServ-enabled network. However, to envision a deployment of such technology in a carrier-grade context would mean significant further effort. In particular, premium paid-for services with SLA (Service Level Agreement) contracts such as targeted by IMS (Poikselka and Georg, 2009)networks would require additional mechanisms able to provide some degree of monitoring in order to asset the SLAs, while IMS by itself does not provide such mechanisms.

The eTOM (enhanced Telecom Operations Map) (Creaner and Reilly, 2005) functional framework is a widespread reference used to model and analyze networks and services activity. From an eTOM point of view, one could argue that IMS does indeed cover the Fulfilment part of service management, but lacks any means to carry out service Assurance. The eTOM framework proposes a complete set of hierarchically layered processes describing all operator activities in a standard way. It is furthermore sustained by a parallel specification of a standard information model, the SID (Shared Information Data) (TMF GB926 Release 4, 2004). It has to be noted however that both tools, the eTOM and the SID, are generic. Also, the eTOM has been designed at times when Services were viewed as centrally controlled and managed, whereas the IMS is really a distributed layer network.

The work presented in this contribution is an attempt to achieve Assurance functionality for QoS-enhanced IMS services following strictly the eTOM specification, thus filling the functional gap as analyzed earlier; furthermore, two architectures are proposed to be compared: a centralized one and a distributed one.

2. IMS and service provisioning

The composition of the supply chain in NGN network is classically described with three layers. The access layer provides IP (v4 or v6) connectivity regardless of the access technologies (Wireless or Wire-line). The service layer therefore supports technology-agnostic

services that are developed independently. The core layer i.e. the control layer is the IMS system which provides the complex signalling responsible for routing sessions between users, invocating services and security-related tasks (Figure 1). The information processing and management are carried out by nodes called CSCF (**C***all* **S***tate* **C***ontrol* **F***unction*) and HSS (***H****ome* ***S****ubscriber* ***S****erver*). The IMS system introduces a control environment similar to the CS session (***S****witched* ***C****ommutation*) but in CP (***P****acket* ***C****ommutation*).

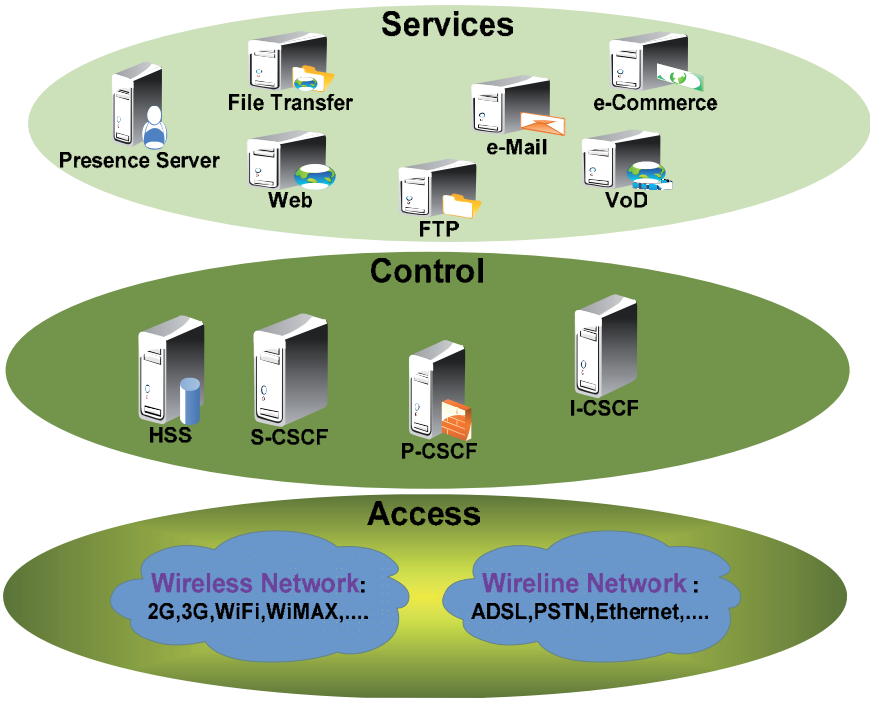


Fig. 1. IMS Layers: Access, Control and Service.

In addition to access unification and diversity of services, IMS introduced a flexible and capable QoS management architecture which organizes exchanges of QoS-related requirements between the control and access layers, allowing resource reservation mechanisms to offer best conditions of supply for e.g. multimedia services.

The service provisioning mechanism of IMS includes three consecutive steps impacting resources: Reservation, Activation and Release (Figure 2).

When a user requests an IMS multimedia service by SIP signalling(Rosenberg et al., 2002) through its attached P-CSCF, the P-CSCF, before forwarding this request must ensure resources availability; this verification is performed through the exchange of Diameter (Korhonen et al., 2010) messages during all media negotiation stages between the two ends (User and AS). An agreement between the client and server can finally lead to change the resource status from reserved into activated. Naturally the PCEF (Policy and Charging Enforcement Function) (3GPP TS 29.210, 2006) applies the relevant QoS policy related to the types of access and transport layers; the most used models are DiffServ(Blake et al., 1998), RSVP(Wroclawski J., 1997) and MPLS (Le Faucheur et al).

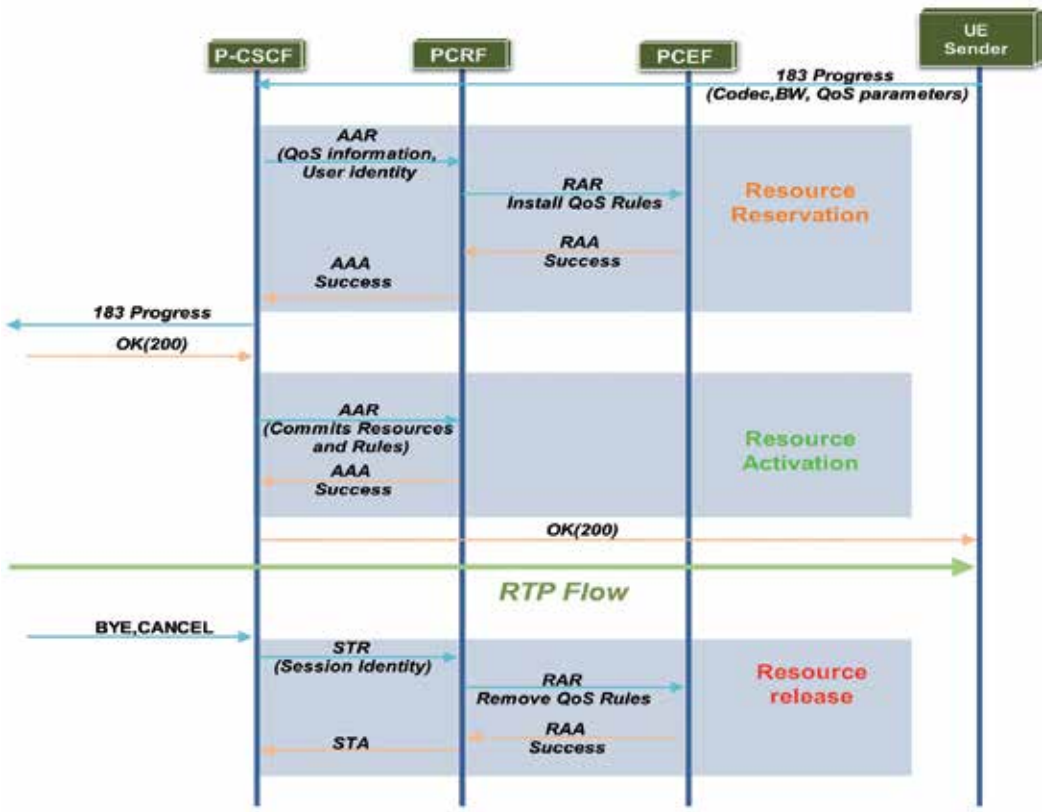


Fig. 2. Service request and negotiation in IMS network with QoS management.

The resources release is carried out at each end of session; the P-CSCF must announce to the PCRF (Policy and Charging Rules Function) (3GPP TR 23.803, 2005) the end of the multimedia session, and the PCRF notifies the PCEF in order to release reserved resources for other applications. QoS management in IMS is a quite flexible on demand mechanism.

3. eTOM (enhanced Telecom Operations Map) architecture

The eTOM is a framework proposed by the TeleManagement Forum and provides a standardized telecom-oriented Business Process map covering all functions of an operator, including service integration and supply. The decomposition layers and functional areas (Customer Service, Resource, and Enterprise) allow detailed operation analysis and to develop solutions according to a well-defined environment. The eTOM has been standardized by the ITU-T (TeleManagement Forum GB921 D, 2010).

The eTOM in its operational part has three main areas: Fulfilment, Assurance and Billing. This section will present only processes related to Assurance, and insist on execution scenarios of SLA (Service Level Agreement)-enhanced services.

3.1 eTOM processes

The 'Operations' area is the traditional heart of the business or service provider (SP). It includes all processes that support client (and network) operations and management. It includes a combination of processes and actions of customer support, including management, provisioning and relationships with partners (Figure 3). The horizontal and vertical processes groupings constitute a matrix formed by a crossing of several processes from level 2, many being derivatives of TOM, which are connected to customer and support operations (FAB).

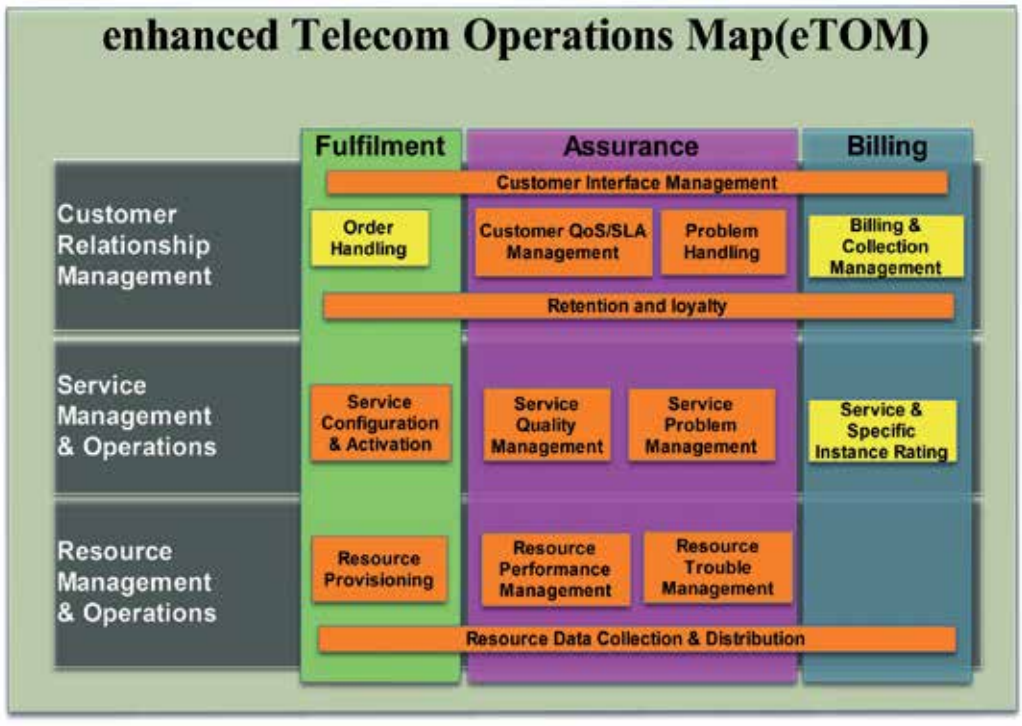


Fig. 3. Operation area in eTOM framework.

A more detailed view of the eTOM business process model (ITU-T Recommendation M.3050.3, 2004) shows a grouping of vertical processes called the FAB columns. These processes are necessary to support operations dedicated to customer satisfaction and operator management:

- **Fulfilment**: Vertical grouping of E2E processes which provide requested services timely and accurately to customers. It reflects business activity. The processes inform customers of their order status, ensure completion on time and customer satisfaction.
- **Assurance**: A group of vertical E2E processes is responsible for implementation of proactive and reactive activities of maintenance to ensure that services are always available and delivered correctly with respect to the SLA. The processes continuously monitor resources status and performance in a proactive way to detect possible defects. They collect performance data and analysis to identify potential problems. In case of

trouble or SLA violation, relevant processes are activated to inform the client about service and trouble status, and to attempt restoration or repair.

- **Billing**: This grouping of vertical E2E process is responsible for collection of appropriate user records, and production of accurate and timely bills, to provide information on resources and services used for payment processing of the customer. In addition, it handles requests from clients on billing, indicates billing status and investigation, and is also responsible for resolving billing issues with respect to customer satisfaction. These processes also support processing of services prepayment.

In addition to the FAB process columns, the Operation area proposes horizontal process groupings:

- **Customer Relationship Management (CRM):** this group of processes supports knowledge of customer needs and includes all necessary features for acquisition, improvement and maintenance of a relationship with a client. It focuses on service and support, and also on retention management, cross-selling, up-selling and direct marketing. CRM also collect customer and applications information, and customization of service delivery to customers. The processes are responsible for identifying opportunities to increase customer value in company. CRM applies to traditional interactions between client and enterprise.
- **Service Management & Operations (SM&O):** This group is focusing on services (access, connectivity, content, solution, composition, etc.). It includes all necessary features for management and operations of communications and information required by or proposed to customers. The focus is on service delivery and management of network and information technology. Some functions involve short-term capacity planning service for a service instance, applying a service design to specific customers or managing service improvement initiatives. These functions are closely related to actual experience of customer. The processes in this group are responsible to meet, at a minimum, QoS goals including performance processes and customer satisfaction with service levels and service costs.
- **Resource Management & Operations (RM&O):** This processes group maintains knowledge of network-related resources (applications, logical and physical infrastructure, communication, management etc.). This group is responsible for managing all these resources (e.g. networks, computer systems, servers, routers, etc.). It is used to provide and support services required by or proposed to customers. The group also contains all features responsible for direct management of these resources (network elements, routers, servers, etc.) used in business process inside operator. These processes are responsible for ensuring that network infrastructure supports an E2E services provisioning. The processes ensure that infrastructure works perfectly, and is available on services and needs and managers.

  The R&O group also has a function that allows collection of information from various sources (e.g. network elements (NE) and/or management systems elements (EMS)), and integrates, correlates and in many cases, summarizes data to be transmitted as information relevant to the service management system. This group also includes processes involved in traditional management of network elements (NEM), because these processes are actually essential elements of any process of resource management. RM&O processes thus manage the network service provider and overall infrastructure to ensure reliable interaction with other service providers.

- **Supplier/Partner Relationship Management (S/PRM)**: This process group supports all FAB business processes: Fulfilment, Assurance and Billing. The processes include issuing requisitions and monitoring them until delivery, mediation of requests that must conform to external processes, validating billing and authorizing payment, as well as management quality of suppliers and partners. When an operator sells its products to a partner or supplier, this is done through the CRM business processes, acting on behalf of the supplier or enterprise in such cases.

3.2 System Information & Data (SID)

Naturally the exchange of information between processes is crucial in the eTOM. The detailed specification of the information supporting the eTOM is provided by the SID informational framework (Figure 4). The SID provides an information model capable of interpreting dynamic and static information of business processes and respects the decomposition of the eTOM. The SID specification uses extensively UML class diagrams.

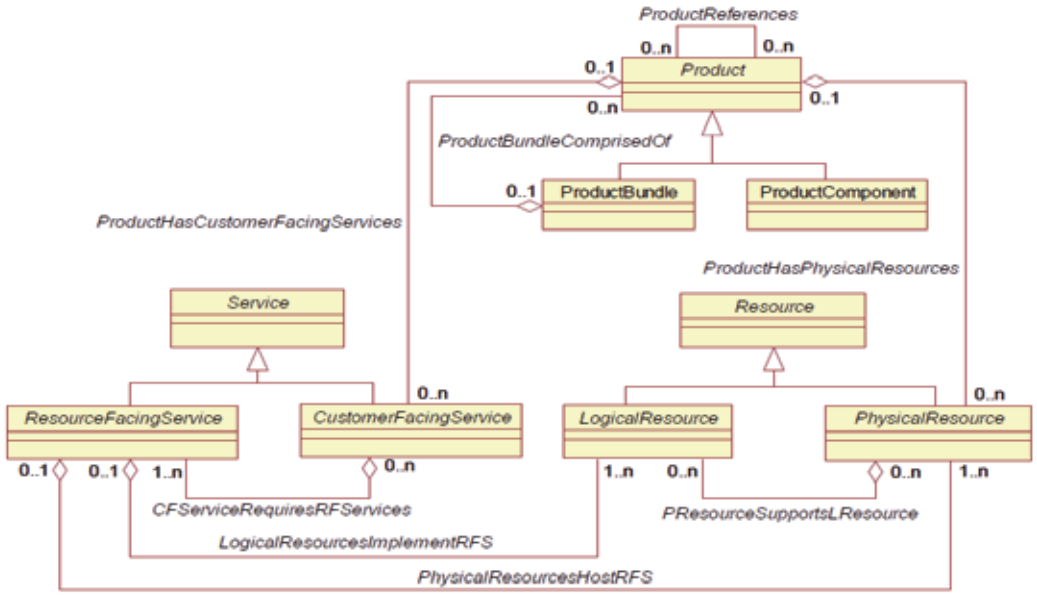


Fig. 4. The main classes of SID: Product, Resource and Service.

- **Product** (TMF GB926 Release 4 Addendum 3 - Product, 2004): in the SID a Product is considered as involving Services and Resources.
- **Service** SID (TMF GB926 Release 4 Addendum 4 SO, 2004): the Product by design is a single or composite service. The Service interacts with the Product to determine its business characteristics, such as customer class and type of service with class *CustomerFacingService*. And the *ResourceFacingService* (TMF GB926 Release 4 Addendum 4 S-QoS, 2004) class exposes resource behaviour for service delivery and its composition for service delivery.
- **Resource**: is divided into two main classes: the **LogicalResource(**TMF GB926 Release 4 Addendum 5 LR, 2004**)** that exposes logical components and services that are necessary

for service design and product needs; and the **PhysicalResource(**TMF GB926 Release 4 Addendum 5 PR, 2004**)** which represents physical components of the network such as routers.

3.3 Execution workflows in the eTOM

The eTOM flows during execution scenarios of SLA-monitored service deliveries describe interactions between business processes as well as the information messages that are exchanged in order to handle both cases: the normal execution and the SLA violation.

3.3.1 Normal execution

The normal execution is a normal state of service delivery without SLA violation and the customer will be billed according to services offered and resources reserved. The operation activates a set of processes and many messages are exchanged between them; the SLA verification requires a mapping between Key Performance Indicators (KPIs) and Key Quality Indicators (KQI) related to service and resource instances.

The SLA verification activates a number of separate processes (Figure 5) which are able to assess QoS according to their positions in the different layers: Customer, Service and Resource.

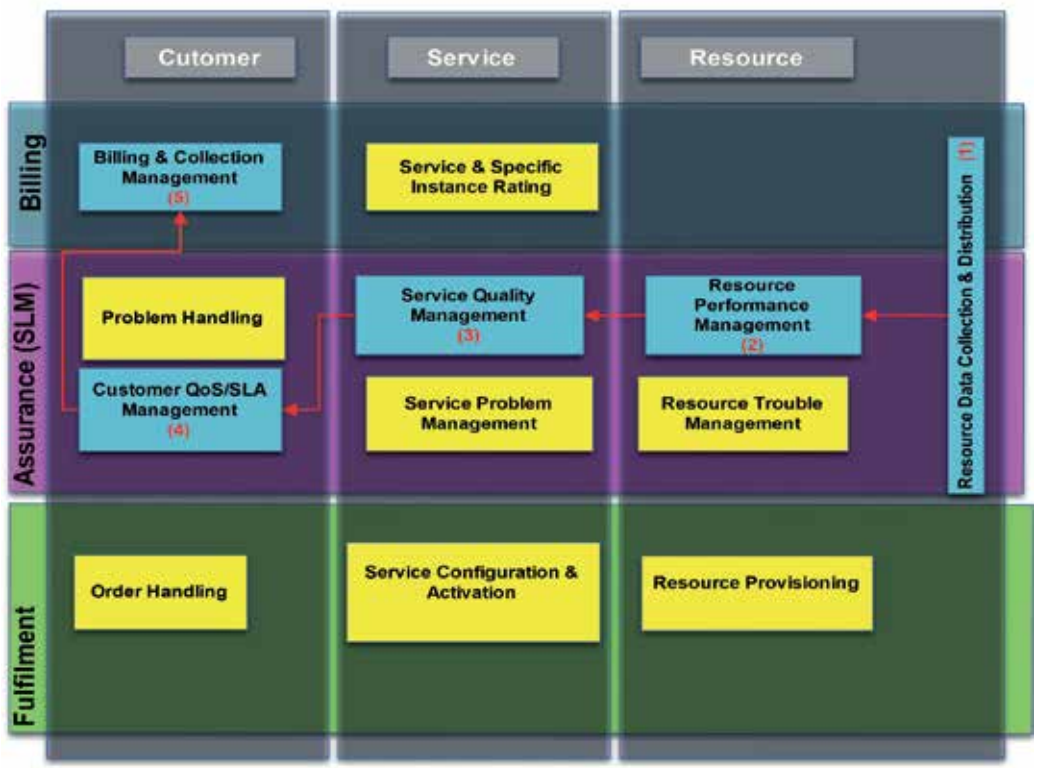


Fig. 5. Active processes in SLA verification.

The SLA verification involves following processes:

- **Resource Data Collection & Distribution**: this process is responsible for the collection of indicators and performance data by contacting all resource agents that provide monitoring, configuration and performance data. The process is also responsible for collecting performance indicators (KPIs) and metrics for all services running in the network. It allows furthermore redistribution of performance data to other processes after aggregation and structuring.
- **Resource Performance Management**: this process reports collected KPIs after filtering and aggregation. The reports provide a structured view of KPIs and a preliminary detection of exceeded thresholds.
- **Service Quality Management**: this process performs a mapping between KPIs and KQIs; it identifies for each service its quality indicators (KQIs) before determining appropriate actions to be performed to calculate them. KQIs values are used to identify failures causes of QoS degradation such a resources failure or lack of capacity in SLA violation.
- **Customer QoS/SLA Management**: is responsible for checking SLA thresholds against measured QoS. After retrieving the KQIs from the Service Quality Management processes and receiving a preliminary report, the process imports the customer profile and SLA parameters to identify thresholds for comparison. It also manages reports of management systems and provides a comprehensive report on the service (metrics, KQIs, key performance indicators, resource use, etc. ...).

The workflow of the SLA verification consists of following steps:

1. When a client requests an IMS service (eg video streaming VoD), the provisioning or "ordering" operation activates all agents in the network to monitor performance indicators and retrieve their values in log files.
2. Resource Data Collection & Distribution retrieves KPIs and metrics collected from different entities in the network. Afterwards, it communicates with the RPM (Resource Performance Management) to identify the existence of critical values and generate performance reports.
3. The performance indicators KPIs collected are sent as XML to Service Quality Management, which identifies indicators KQIs and realize mapping function, and comparing with thresholds are specific to requested service.
4. The Customer QoS / SLA Management uses the loaded profile of customer to identify product thresholds to apply to data collected prior to drafting of audit report of SLA against QoS.
5. The process Billing & Collection Management performs charging functions and taxation with received information to make bills.

3.3.2 SLA violation

The SLA violation scenario begins with a simple verification as above, but in this case a threshold violation occurs. In this case the eTOM provides an escalation mechanism: first, the Resource Layer attempts to solve the problem locally, while warning the Service Layer in order to plan alternative solutions. If the trouble persists, Service processes must perform an alternative service configuration produced by an Ordering operation; this new configuration may be the best solution and is followed by a return to normal SLA verification. The operation chronology consists of three stages: detection and attempted correction, reconfiguration, return to normal verification of SLA.

A real-time continuous monitoring of provided services allows early alerts concerning exceeded thresholds and resource failure alarms, which are main causes of violations and SLA unconformity. Most interactions occur within Assurance processes, but interactions are also concerning the Fulfilment processes, and violation is considered for reimbursement through the Billing processes.

Two specific processes handle the escalation mechanism depicted above: the *Service problem Management* and *Resource Trouble Management* processes (Figure 6).

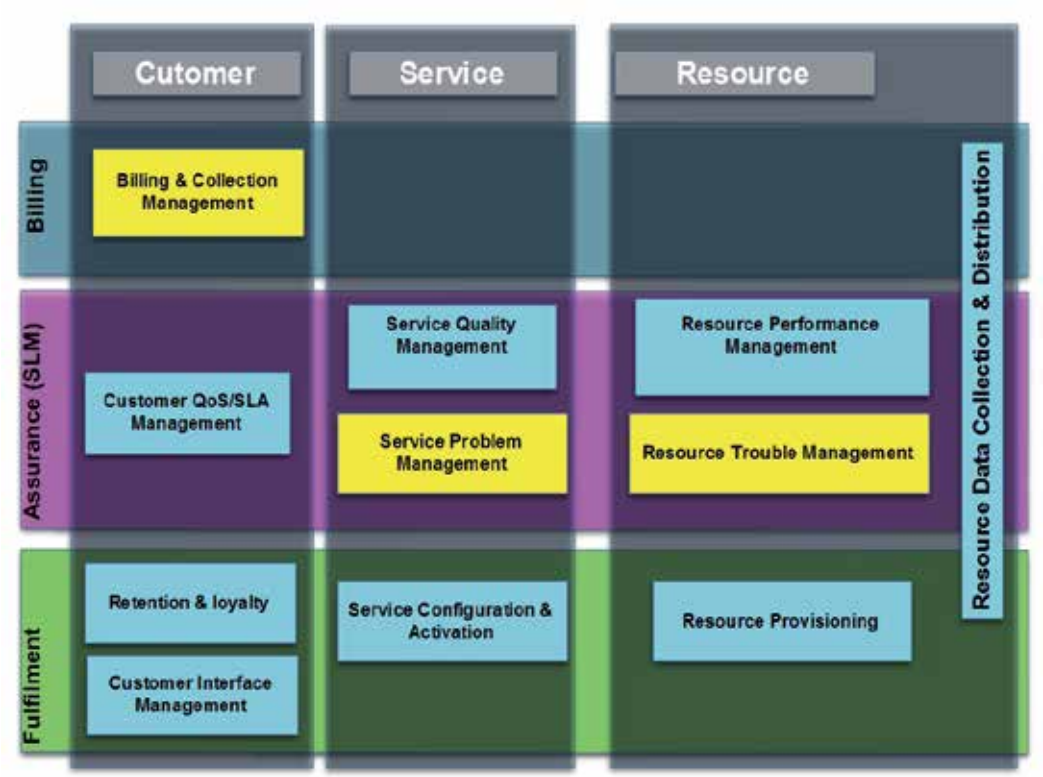


Fig. 6. Active processes in SLA violation.

The goal of these two processes is to perform a restoration of services and resources in short time, and to locate troubles before their expansion, with an optional notification to the user.

The operation is initiated by a usual collection of data by the **RDC&P** process when detecting and exceeded threshold. The process sends relevant information to **RPM** to alert the **RTM** process; in case of a component failure the communication is done directly between **RDC&P** and **RPM**.

The RPM process sends details to the Service Quality Management (SQM) and to the RTM process, depending on the type of trouble, trying to start procedures for resource restoration; for each attempt it notifies the Service Problem Management (SPM) process to synchronize their information about troubles (Figure 7).

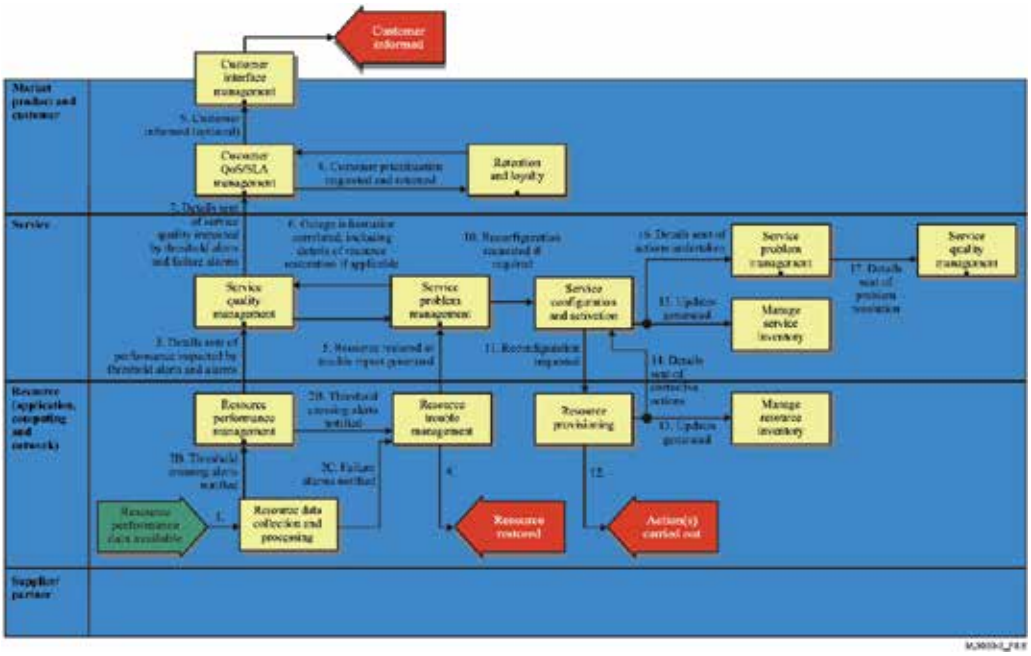


Fig. 7. Processes Flow in SLA execution with violation.

The communication between **SPM** and **SQM** aims to correlate their information about troubles whether solved or not; the **SQM** process sends details of impact on services and alarms to **Customer QoS / SLA Management** (**CQoS / SLAM**), a process that specifies SLA violation and customer importance by obtaining all this information from **Retention & Loyalty**, and finally notifies the customer about QoS degradation according to its importance.

If the cooperation between the Service Problem Management and the Resource Management Trouble processes is unsuccessful, the SC&A process will be activated to perform its own corrective action, such as a new configuration. The new configuration will take into account all resource constraints and infrastructure development and service contract terms.

The reconfiguration proposed by SC&A follows exactly the steps of the Ordering operation, and is finalized by launching normal SLA verification, and tries to close all open troubles reports in SPM and RTM. The CQoS / SLAM process can inform the customer about service restoration and quality with the possibility of sending a QoS report.

4. Issues

3GPP specifications provide a basic QoS management architecture for the IMS network which ensures an adequate level of service compared to best effort service. However, the IMS services need to be monitored and managed by a set of mechanisms and methods taking into consideration constraints of the business enterprise. Such a set is explicitly proposed by the eTOM. The eTOM describes its operations and processes in ways that are generic and applicable to any transaction and promises to be fully applicable to the IMS architecture with no applicability constraints. The next step of the study is therefore to plan a mapping strategy in order to map eTOM processes to IMS functions.

5. Functional architecture

A first step in this undertaking is to match IMS functionality with eTOM processes. The resulting set has furthermore to be enriched by eTOM processes relevant to Assurance and Fulfilment. This broader set forms the basis to select different SID entities necessary to carry out these processes. The SLA execution procedure as defined in eTOM model requires the cooperation of several processes belonging to Assurance and Fulfilment of the 'FAB' area, and spanning the three business layers: Customer, Resource, and Service. These eTOM processes will be activated sequentially (Figure 8).

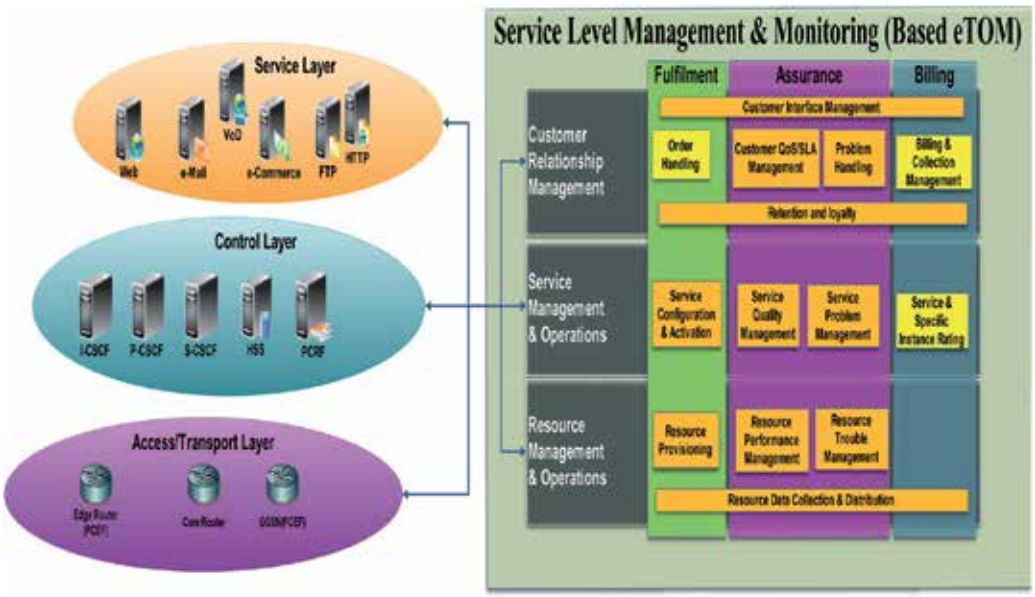


Fig. 8. eTOM and IMS interactions.

The processes belonging to the Assurance layer correspond to the monitoring aspect of this operation, related to Fulfilment for restoration and supply. In order to link eTOM processes to the IMS network, a new component entitled Monitoring, Configuration, Data Collection is required, which clusters the core modules to communicate with these entities.

In the IMS network, the diversity of entities and their various communication protocols require multi-protocol components which can implement all the necessary monitoring and correction operations. An additional constraint is that performance data collection and detection of services should be executed in real time or near real.

5.1 Design

The WSOA (Web Service Oriented Architecture) appears as a valid choice for such a distributed system. The SOA (Mark and Hansen, 2007) concepts will allow to implement EJB (Rima, Gerald, and Micah, 2006) based SOA modules supporting the processes of each component, exposing web services communications via XML/SOAP (Simple Object Access Protocol) /HTTP (Newcomer E., 2002). Three SOA modules have been designed, each of which supporting a part of the targeted eTOM business processes and their associated SID

entities. In addition, a BPEL (Business Process Execution Language) (Poornachandra, Matjaz, and Benny, 2006) component has been designed to orchestrate the various processes and to organize the desired operations (Figure 9).

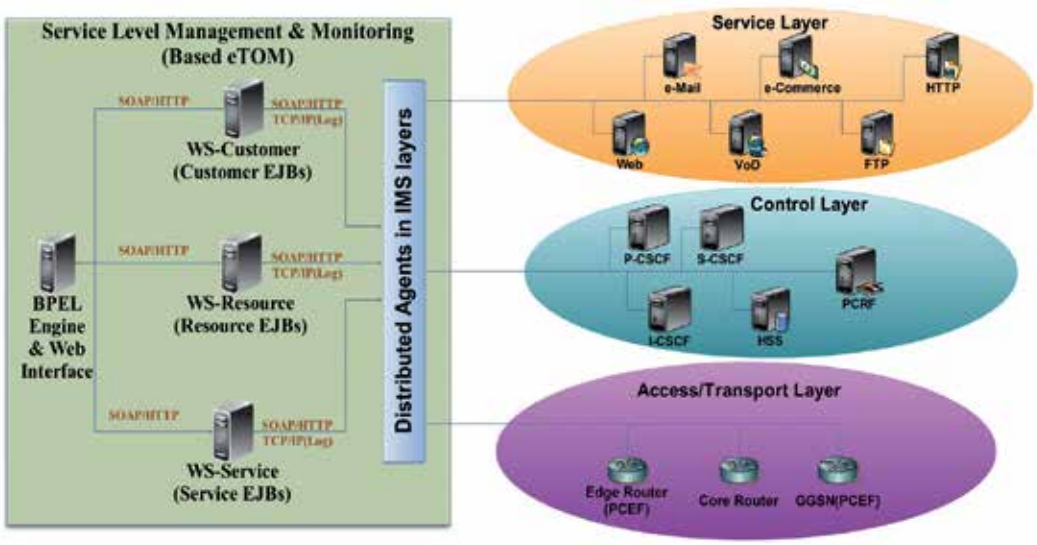


Fig. 9. Implementation architecture.

The three modules of the monitoring system are: WS-Resource, WS-Service, and WS-Customer; each one exposes a set of web services specified using WSDL (Web Services Description Language) (W3C Recommendation, 2007). These web services are invoked and synchronized by the central BPEL component that provides moreover tools such as a web interface that tracks performance of overall network, SLA operation, processes execution and monitoring of physical and logical network resources (Figure 9). The SLM&M (Service Level Monitoring and Management) architecture contain:

- Translation Business Processes: EJB (Enterprise Java Bean) for represent each processes, its functions and information processing.
- Presentation of WSs: WSDL (Web Service Description Language).
- Processes Communication: XML/SOAP (Simple Object Access Protocol).
- Operations Orchestration: BPEL (Business Process Execution Language).
- Communication between SLM&M and the IMS network entities: TCP/IP, XML.

5.2 Centralized architecture

The initial architecture is centralized and enables a selective monitoring of consecutive operations related to SLA and it verification. This system allowed demonstrating the steps of the verification operations, the different KPIs and KQIs of service, and some operational limitations (Raouyane B. et al., 2011).

In order to simplify SLM&M, the number of exposed web services has been limited to eTOM level 3 business processes. Naturally processes of level 4 are implemented via appropriate methods within web services.

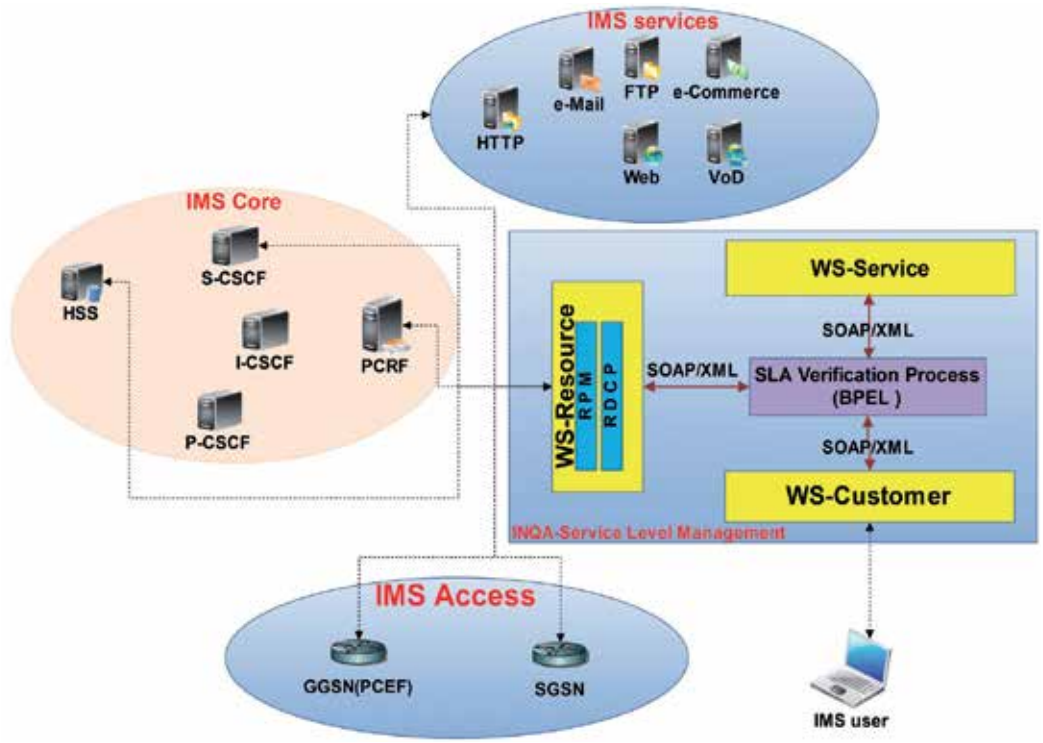


Fig. 10. SLM&M Centralized architecture.

Before analyzing the different SOA modules, it is useful to introduce the interfaces between SOA modules and network. These interface agents take (Figure 10) in charge low-level detections and calculations and transmit their results to the SOA modules via dedicated socket interfaces:

- The IMS agent scans S-CSCF activity and detects service launching ;
- The Application Server agent scans Application Server activity in order to identify the customer parameters ;
- The Router agents perform network analysis tasks in order to calculate KPIs that will be transmitted to SOA modules

The SOA Modules expose each eTOM layer:

- WS-Resource: This SOA module is composed of classes implementing operations defined in the Resource layer of eTOM, as well as corresponding SID entities. It implements two main eTOM processes already discussed in functional architecture: Resource Data Collection & Processing, and Resource Performance Management. Both of them are exposed as web services.
- WS-Service: This module implements various SID entities and operations defined in the Service layer of eTOM. The module exposes processes as web services responsible for quality indicator mapping and analyzing.
- WS-Customer: This module implements functionality defined in the Customer layer of the eTOM and its SID model. It exposes Customer QoS/SLA Management web service

responsible for SLA verification. It retrieves the KQIs of the currently delivered service, loads the customer profile and subsequently detects any SLA violation.

- BPEL Engine: The BPEL Engine module implements a BPEL process that invokes the web services described above and synchronize their interaction.
- Web interface : To monitor the SLA operation and its process, the BPEL engine features a web interface that allows to:
  - Show messages exchanged between web services (XML/SOAP) and modules.
  - List performance indicators collected from the network layer entity
  - Monitor the activity and performance of physical resources such as network routers and logical entities such as CSCFs and the HSS (Home Subscribe Server).
  - Check the provisioning chain of QoS management: PCRF, PCEF.
  - View the results of the audit and SLA verification, the customer class, and values of KQIs.

5.3 Distributed architecture and continuous monitoring

To reduce SLM&M complexity, an enhanced architecture proposes to split the RDC&P into many smaller distributed and decentralized components. Additionally, continuous monitoring functionality has been added to the system (Figure 11).

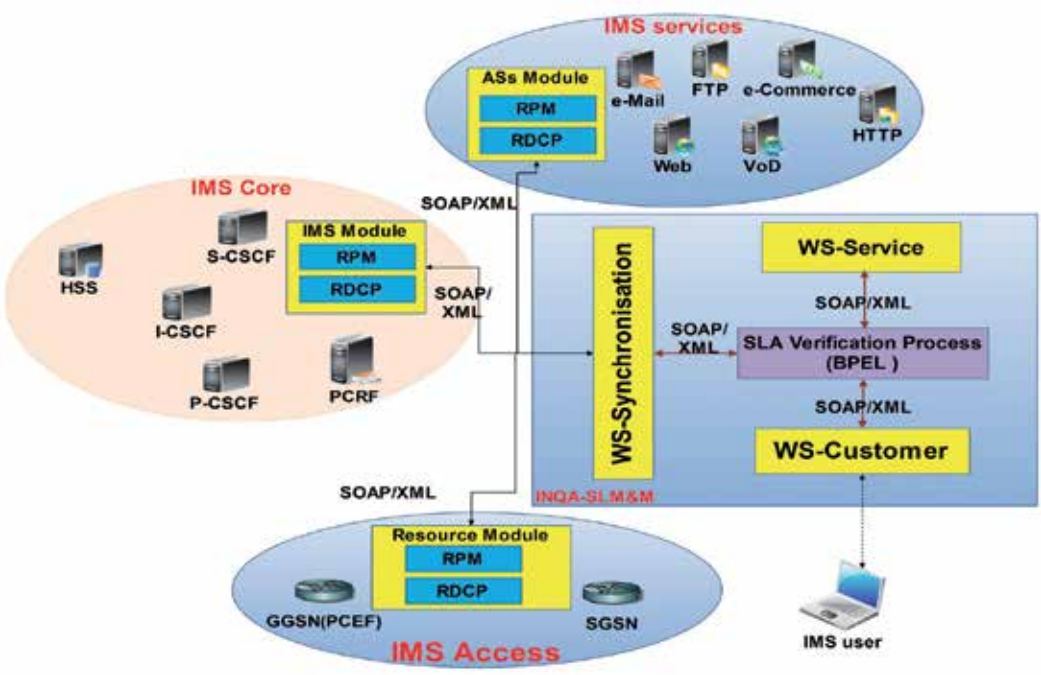


Fig. 11. SLM&M distributed architecture.

However, the centralized web services that allow KPIs and KQIs monitoring are still relying on BPEL technology and still are very resource and time consuming. Indeed, the distribution of the Resource processes allows not only to share processes of KPIs but also to evaluate services locally. Thus, the distribution of EJB modules becomes necessary to

incorporate mechanisms for monitoring locally but also to allow a local correction of QoS degradation and anomalies.

The new functional architecture of SLM&M (Figure 12) consists therefore of two main modules:

- **Assurance Layer***:* represents the SLA verification process as defined in eTOM in both layer Customer and Service. Thus, processes that are related to operations Fulfilment and Assurance, and the information and data is stored in Customer Inventory and Service Inventory.
- Monitoring Layer:
  - Is distributed, and contains a set of agents and probes that are able to recover all data in real time (signalization, logging, reservation, configuration, policy, routers status, etc. ..) and implements all Resource layer processes for SLA verification: Resource Data Collection & Processing (RDC&P) and Resource Performance Management (RPM), which are related to each IMS layer (Access, Control, Service).
  - Contains a set of processes that are functional in ordering and other SLA operations of WS-Resource. Also, a synchronization module that is necessary for detection and control of events in the network, such as planning activities and communications on one hand between the distributed modules (first part) and also between Web services exposed (Layer).

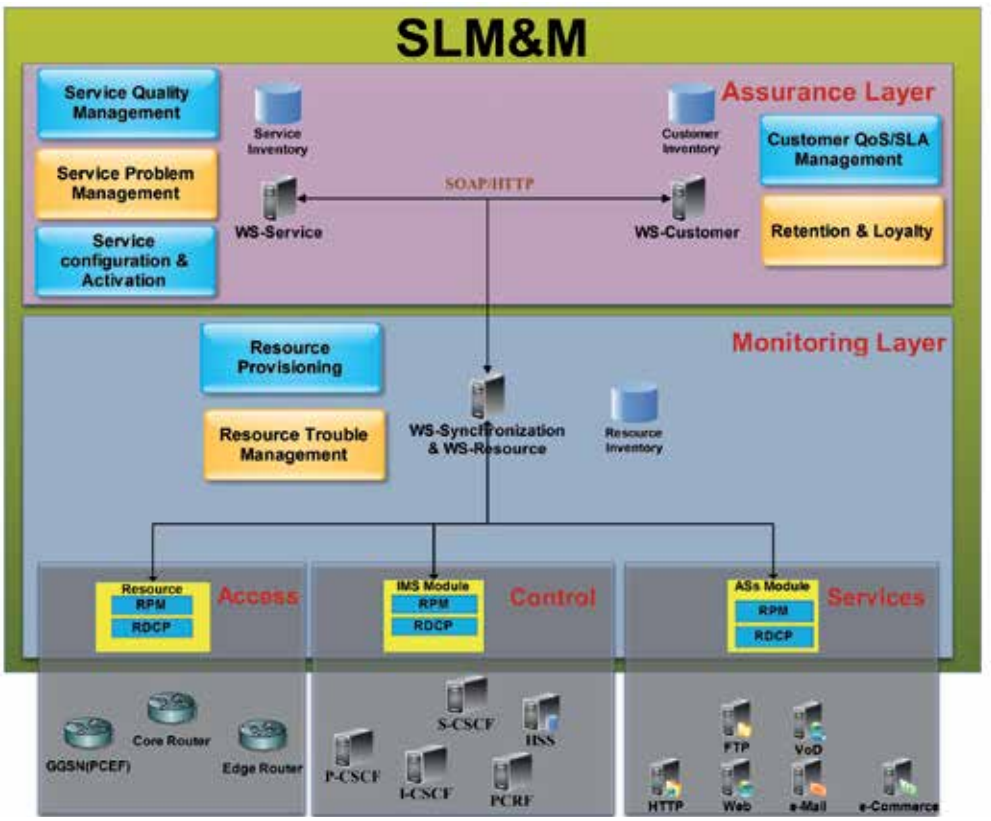


Fig. 12. The main layer of SLM&M: Monitoring, Assurance.

The proposed functional architecture supports three communication channels between different modules:

- TCP / IP between agents and the synchronization module,
- SOAP / XML between layers of eTOM (Resource, Service, Customer) or WSs
- With ability to use XMLCONF between the synchronization module and management in case of SLA violation.

5.4 Correction architecture

The architecture of SLM& M consists of two layers: Assurance and Monitoring, by analogy with the previous architecture (distributed). The Monitoring layer is distributed and contains only two main processes of collection and processing of information.

New processes must be integrated in a centralized way; a distributed integration can overload collection agents in routers and. For example in a router memory is crucial,; collection agents and processes DRC & P and RPM are reasonable for just performance collection and data local treatment. However, the addition of another process could overload the router that needs its capacities for traffic conditioning and processing.

The Resource Trouble Management process (RTM) catches alarms that reflect a degradation of service resulting from a physical or logical related to equipment; this process then tries to make a preliminary correction of the service and notifies WS-Service.

The WS-synchronization process is located in the same server as WS-Resource, so that this server can synchronize incoming events and data collection, and decide either to perform a normal SLA verification, or to report a violation. Also, the Resource Provisioning process is responsible for making resource reservations with respect to solution recommendations provided by WS-Service. WS-Service adds to its repertoire Service Problem Management (SPM) processes.

The interaction between WS components of SLM&M is through SOAP/HTTP, whereas the interaction between Monitoring layer of SLM &M and IMS layers uses Java-based client / server communication, with a spare possibility of using XML/RPC (Mi-Jung et al., 2004) between the (Resource, IMS ASs) and WS-Resource modules.

6. Implementation and results

The implementation encompasses three fundamental components:

- The IMS network for service delivery: control entities (CSCF, HSS) and Application server for video streaming VoD.
- The QoS management: PCEF and PCRF.
- The Monitoring and managing System: SLM&M.

6.1 Trial infrastructure in SLA verification

The trial architecture exposes the function of each component.

The test bed is composed of (Figure 13):

- A core router and two edge routers (Linux boxes) defining a DiffServ-enabled network on which are connected an IMS terminal ad an Application Server;
- This network is controlled by the OpenIMS (Open IMS Core) system which is deployed in the core router Linux box;
- A management server supports the QoS monitoring/Assurance functionality.

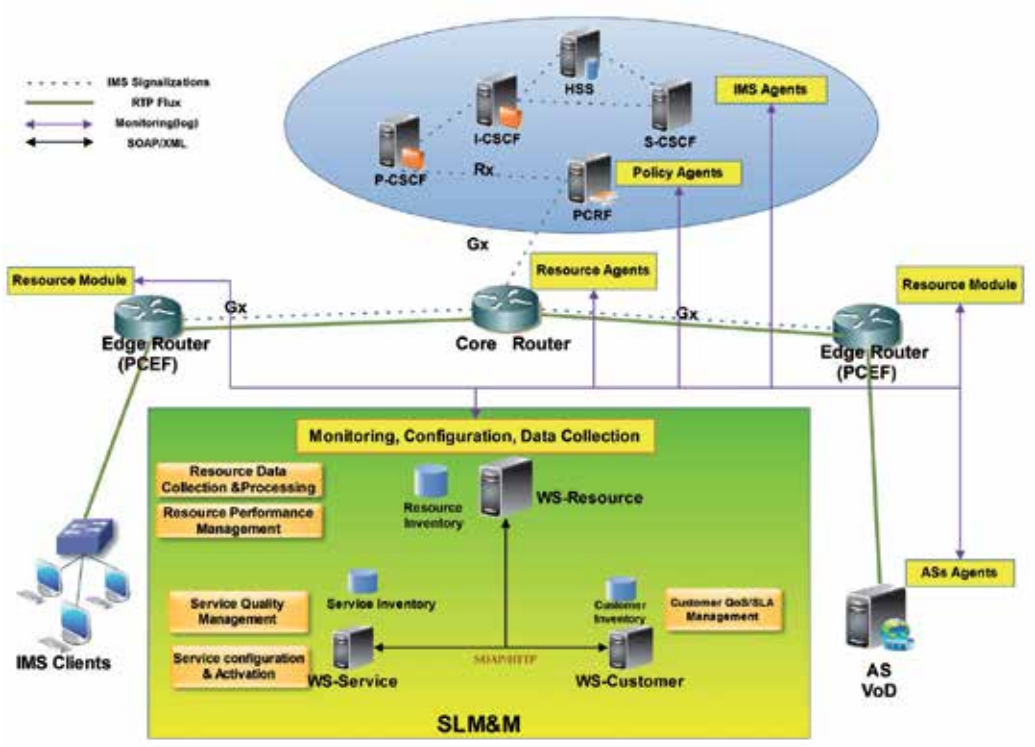


Fig. 13. Centralized trial infrastructure.

6.2 Trial infrastructure for SLA violation

The implementation architecture features two sub-architectures for the service provisioning and for service management and correction services.

The Supply Architecture which contains an IMS network that includes both the signalling and the media planes. The architecture includes three routers to transmit the media stream; a central router supports the IMS system. The PCRF (Policy and Charging Rule Function) is becoming an autonomous entity and includes other features such as policy management, and both edge routers include the PCEF (Policy and Charging Enforcement Function) functionality to receive and execute policies or PCC rules (Figure 14).

Monitoring and management architecture: SLM & M is divided into two layers

- *Monitoring Layer*: contain the two WSs Resource and Synchronization, with the integration of RP and RTM processes and Resource Inventory, so the layer includes the functionality of PCRF for QoS management and control.

- *Assurance Layer*: contain both servers and WS-Customer and WS-Service, and integrate process and Fulfilment and Assurance, that will be activated in SLA correction and violation.

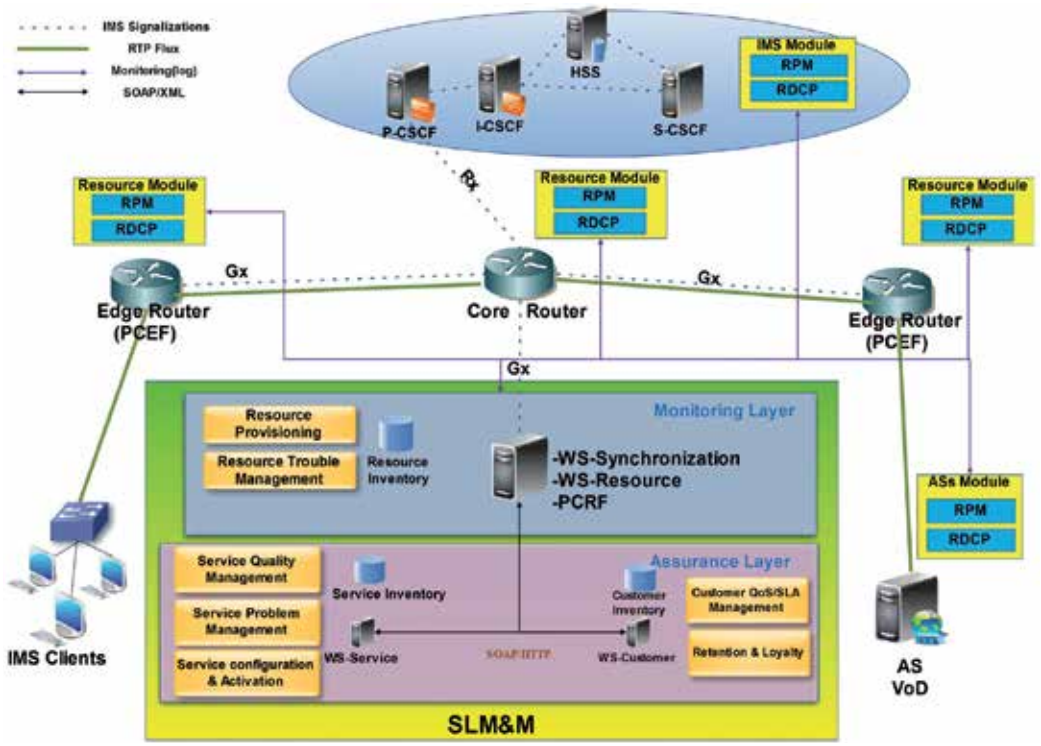


Fig. 14. Centralized trial infrastructure.

The supply architecture provides a set of IMS services, when a client requests a VoD streaming service, the provisioning chain stimulates IMS entities to provide a resource reservation and QoS management. The SLM&M in supply, after ordering operation, start collection of configuration data for a normal SLA execution. Anomaly detection or exceeding threshold causes an activation of SLA violation processes for restoring service into normal level.

The SLM&M must be reactive by rapid detection of QoS degradation or anomalies, followed by an attempt to resolve troubles, that activates the Assurance process and if necessary the Fulfilment process.

6.3 Scenarios

The test scenario includes three cases, a customer Alice with Platinum class that requires a VoD service:

- The client receives service with perfect QoS;
- An overloaded network with a slight QoS degradation;
- Network Congestion and violation of SLA.

6.4 Results

The results expose several parameters relating to monitoring service and performance of SLM&M in centralized and distributed architecture, and response time in trouble detection and SLA violation.

6.4.1 SLA verification in centralized architecture:

Case 1: The QoS offered to Alice and Bob matches SLA contract, perceived video quality is satisfying (Figure 15).

Fig. 15. Video bandwidth =128kbps.

Case 2: the network conditions, hence video quality, deteriorate proportionally to mass of competing services for lost packets and reduced flow rate (Figure 16).

Fig. 16. Video bandwidth =76kbps.

Case 3: competing services overload the routers: queues fill in gateways, impacting delay and jitter. Routers discard packets in excess, this causes static pixels in video (Figure 17) and in some cases service cancellation.

Fig. 17. Video bandwidth =40kbps.

The platform succeeds in identifying accurately deterioration of delivered services. The cost in terms of response time has been evaluated as well. It is observed that response time for Resource-WS is much longer than for other web services, due to complexity of its tasks (Figure9, 10).

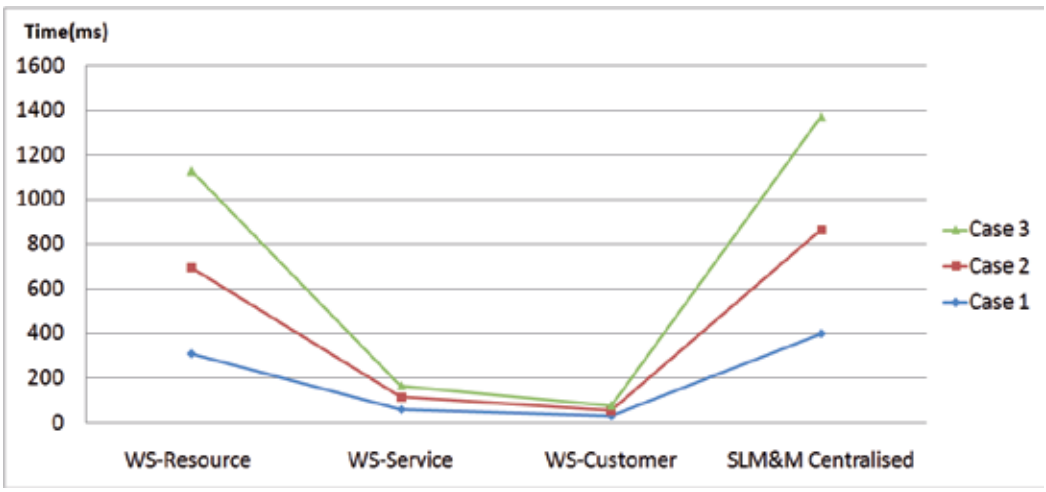


Fig. 18. Response time for different web services (Client: Alice).

The number of web services and their internal functionality has a considerable impact on running time of SLA verification. This led to limit the exposed eTOM processes to level 3 and to implement sub processes via internal java methods.

6.4.2 Centralized vs distributed architecture

The execution time in SLA verification is composite, and is directly related to processing time in each WS. This time varies depending on the number of planned operations, WS state and SLM&M conception. Similarly, nature of communication technology between entities

plays a vital role in reduction of complexity and processing, which highlights the advantage of using TCP /IP for exchange parameters of service and performance indicators and transmission at higher levels in order to achieve continuous monitoring.

The response time of WS-synchronization and other agents resource is short compared to WS-Resource in SLA verification, because of the processing performed locally in each device, that has a potential to reduce traffic between Monitoring and Assurance layers, as well as it reduces response time compared to centralized architecture in tree cases (Figure 19).

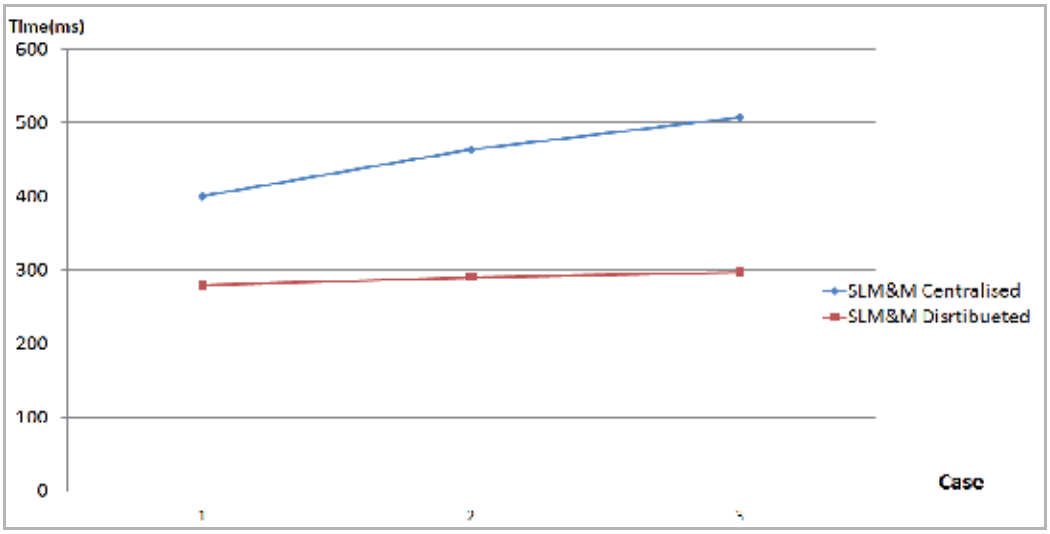


Fig. 19. Response time (ms) for centralized and disturbed architecture.

The processing of parameter flow in entity level allow a real-time control of multiple QoS and services, however a router must have enough memory for traffic conditioning, although using control function as treatment and comparison with thresholds can reduce its capacity in terms of CPU and memory.

The distributed architecture provides a Monitoring layer that alerts Assurance layer within a very short time or near real time, and allows rapid processing of SLA violations, compared to another architecture where several treatments must be executed to detect degradation QoS or failure of an entity.

6.4.3 SLA violation

Alice has registered in the IMS system with QoS classes Gold. The goal is to perform SLA Assurance tests in three representative cases and to compare results for SLA correction with Assurance and Fulfilment.

The MOS-AV (Mean Opinion Score-AudioVisual (BELLAFKIH et al., 2010) is a quality indicator for detecting user satisfaction requiring a video flow such as VoD (Video on demand) or IPTV (IP Television). It is a quantitative indicator taking values between 0 and 5. The MOS-AV reflects user satisfaction and SLA violation in our case the value 2.5 represent a QoS degradation and SLA violation (Figure 20).

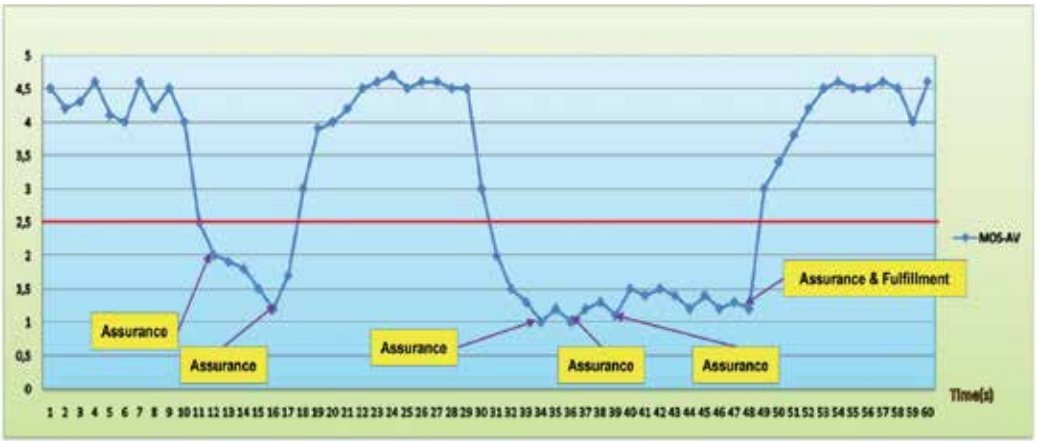


Fig. 20. Customer satisfaction during a time interval after SLA verification and correction.

The MOS-AV indicator reflects customer satisfaction. When detection thresholds are exceeded or values of MOS-V become critical, SLA violation has launched the first attempt with confidence and was successful in restoring normal levels of service after 7 seconds. The second violation that requires intervention of Assurance & Fulfilment takes 17 seconds to restore the service. These results are justified by the architecture that used WSOA and interaction between different WSs and attempted solutions.

7. Conclusion

The proposed approach is based on the QoS provisioning architecture proposed by 3GPP with eTOM Assurance capabilities of QoS monitoring. SLM&M uses the new concepts of SOA and BPEL in managing and monitoring network and also must meet several constraints of instrumentation. The first version of the platform was centralized to address all performance data in a central node, this design offered SLA verification but still remained isolated from IMS network and real events. However the IMS network requires permanent and real time monitoring rather than just a sporadic SLA verification.

The solution to distribute Resource layer and processes of the eTOM and their adjustments to each layer of the IMS (Access, Control, and Service) is appropriate, as the creation of a WS- Synchronization which synchronizes operations process WS-Resource and their agents in network layer, in addition provides operations in the monitoring layer of SLM & M. The distributed architecture demonstrates its ability in terms of response time and is preferable to a centralized SLM&M.

The life cycle of SLM&M has three main stages: a real-time monitoring of services and resources to detect anomalies or degradation, followed by a stage of responsiveness to correct troubles, and the final step is to be proactive in order to estimate the behaviour of service and resource by correlation and root cause analysis of service impact. The proactive property will be integrated with QoS mechanisms that predicted from current data, a mathematical model or stochastic processes that come into the perspective.

8. Acknowledgment

Special thanks are due to MÉDITELECOM operator for its financing, and supporting.

9. References

3rd Generation Partnership Project "Charging rule provisioning over Gx interface (Release 6)", 3GPP TS 29.210 V6.7.0 2006-12, available at http://www.3gpp.org/ftp/Specs/html-info/29210.htm.

3rd Generation Partnership Project, Evolution of policy control and charging (Release 7), 3GPP TR 23.803 V7.0.0 (2005-09), available at http://www.3gpp.org/ftp/Specs/html-info/23203.htm.

Bellafkih, M.; Raouyane, B.; Errais, M.; Ramdani, M.;MOS evaluation for VoD service in an IMS network, *I/V Communications and Mobile Network (ISVC), 2010 5th International Symposium,* Rabat , Morocco.

Blake S., Black D., Carlson M., Davies E., Wang Z., and Weiss W., An Architecture for Differentiated Services, December 1998, RFC 2475, available at http://www.ietf.org/rfc/rfc2475.txt

Creaner M. J. and Reilly, J. P, NGOSS Distilled: The Essential Guide to Next Generation Telecoms Management. The Lean Corporation 2005.

Enhanced Telecom Operations Map (eTOM) The Business Process Framework for the Information and Communications Services Industry, Addendum D: Process Decompositions and Descriptions Release 6.0 GB921 D; TMF.

J. Wroclawski, The Use of RSVP with IETF Integrated Services, September 1997, RFC 2210, available at http://www.ietf.org/rfc/rfc2210.txt

Korhonen, J., Tschofenig, H., Arumaithurai, M. Jones, M., Ed., and A. Lior, "Traffic Classification and Quality of Service (QoS) Attributes for Diameter", RFC 5777, February 2010, available at http://www.ietf.org/rfc/rfc5777.txt

Le Faucheur F., Wu L., Davie B., Davari S., Vaananen P., Krishnan R., Cheval P., Heinanen J., "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, May 2002, available at http://www.ietf.org/rfc/rfc3270.txt.

Mark Hansen, D. (2007) *SOA Using Java Web Services*, Prentice Hall, New Jersey, USA

Mi-Jung Choi, Hyoun-Mi Choi, Hong, J.W., Hong-Taek Ju, XML-based configuration management for IP network devices, *Communications Magazine IEEE*, July 2004, Volume: 42 Issue:7, pages: 84 – 91.

Newcomer, E. (2002) *Understanding Web Services: XML, WSDL, SOAP, and UDDI*, Addison-Wesley Professional.

OpenIMScore – Open source implementation of IMS Call Session Control Functions and Home Subscriber Service (HSS) -http://www.openimscore.org/

Poikselka M. and Georg M. *The IMS: IP Multimedia Concepts and Services*, *John Wiley & Sons* Inc. Chichester, England2009.

Poornachandra, S., Matjaz, J., and Benny, M. (2006) *Business Process Execution Language for Web Services BPEL and BPEL4WS,* 2nd Edition, Paperback.

Raouyane B., Bellafkih M., Ranc D., QoS Management in IMS: DiffServ Model', Paper Presented at the *Third International Conference on Next Generation Mobile Applications, Services and Technologies*, 15-18 September 2009. Cardiff, Wales, UK.

Raouyane B.; Bellafkih M.; Errais M.; Ranc D.; WS-Composite for Management & Monitoring IMS Network; *International Journal of Next-Generation Computing (IJNGC)* - ISSN 2229-4678, eISSN 0976-5034.

Rima, P. S., Gerald, B., and Micah, S. (2006) *Mastering Enterprise JavaBeans 3.0*, Paperback.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M. and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002, available at http://www.ietf.org/rfc/rfc3621.txt

SERIES M: Telecommunications management network Enhanced Telecom Operations Map (eTOM) –Representative process flows, ITU-T Recommendation M.3050.

Shared Information/Data (SID) Model System View Concepts and Principles, *GB926 Version 1.0, Release 4.0*, January 2004.

Shared Information/Data (SID) Model, Addendum 3 - Product Business Entity Definitions GB922 Version 3.1, NGOSS Release 3.5 July 2003.

Shared Information/Data (SID) Model, Addendum 4SO – Service Overview Business Entity Definitions, GB922 Addendum-4SO, NGOSS Release 3.5 July 2003.

Shared Information/Data (SID) Model, Addendum 4S-QoS Quality of Service Business Entity Definitions, GB922 Addendum – 4S-QoS Version 1.0, NGOSS Release 3.5 July 2003.

Shared Information/Data (SID) Model, Addendum 5LR – LogicalResource Business Entity Definitions, GB922 Addendum-5LR Version 1.0, NGOSS Release 3.5 July 2003.

Shared Information/Data (SID) Model, Addendum 5PR – Physical Resource Business Entity Definitions, GB922 Addendum-5PR Version 3.0 NGOSS Release 3.5 July, 2003.

Web Services Description Language (WSDL) Version 2.0, W3C Recommendation 26 June 2007, http://www.w3.org/TR/wsdl.

Part 2

Quality of Services

4

A Testbed About Priority-Based Dynamic Connection Profiles in QoS Wireless Multimedia Networks

A. Toppan, P. Toppan, C. De Castro and O. Andrisano
IEIIT-CNR, National Research Council of Italy & WiLab, University of Bologna, Bologna, Italy

1. Introduction

The ever-growing demand of high-quality broadband connectivity in mobile scenarios, as well as the Digital Divide discrimination, are boosting the development of more and more efficient wireless technologies.

Despite their adaptability and relative small installation costs, wireless networks still lack a full bandwidth availability and are also subject to interference problems.

In context of a Metropolitan Area Network serving a large number of users, a bandwidth increase can turn out to be neither feasible nor justified. In consequence, and in order to meet the needs of multimedia applications, bandwidth optimization techniques were designed and developed, such as Traffic Shaping [1-3], Policy-Based Traffic Management [4-8] and Quality of Service (QoS) [9-17].

In this paper, QoS protocols are adopted and, in particular, priority-based dynamic profiles in a QoS wireless multimedia network. This technique [18-20] allows to asssign different priorities to distinct applications, so as to rearrange service quality in a dynamic way [21,22] and guarantee the desired performance to a given data flow.

In particular, the platform can manage two levels of priority: among different users and within a single user's connection.

In the former kind of priority management, those users whose guaranteed bandwidth is higher, will be proportionally assigned a greater part of the shared bandwidth.

The latter case refers to each single user, whose distinct services are assigned distinct priorities. Each profile, in fact, allows the real time management of services, and the priority parameter is used to queue the desidered services properly.

A complete testbed involving 80 users approximately is here presented, where such technique is adapted to the specific requirements of the plant.

The network infrastructure installation is detailed, the whole QoS system developed is described and four measurement campaigns are reported.

The whole testing was directed by WiLab (www.wilab.org), which includes the IEIIT-CNR (National Research Council of Italy, IEIIT Bologna unit) and a portion of the TLC scientific community at Bologna University (Italy). The design and technical aspects of the problem were and are still being carried out by such group.

The proposed platform aims at supporting a Wireless Internet Service Provider (WISP) in the management of its network infrastructure in a user-friendly and straightforward way. It can be accessed through the Internet and lets the network manager define different access profiles and supervise all the users' connections.

The QoS service, in particular, allows to set each user's minimum bit rate guaranteed and maximum supported, enable services such as VoIP, FTP, Mail and P2P and assign them specific priorities.

The network scenario installed and used for the testbed includes an Internet gateway, a server which hosts the whole infrastructure control system and five sectoral distribution devices.

The software platform allows to define some distinct kinds of priority-based connection profiles, each characterized by a set of different parameters and a diverse commercial value. Personal data can also be managed, each corresponding to the user's kind of connection profile subscribed.

Reports about connections and traffic statistics are also at disposal, also useful to law purposes. A continuous monitoring of the wireless network infrastucture is also possible.

All the above features can be easily managed through specific user-friendly portals.

This kind of services are fundamental in many application fields, ranging from Intelligent Transportation Systems (ITS) and Infomobility to "Smart Cities", where wireless applications guide the user in most of his activities.

The paper is organized as follows: Section 2 describes the testbed setup and the QoS software developed. Section 3 details the results of the four measurements campaigns.

These techniques, addressed to to real applications, are discussed in the following: the PEGASUS project about the support of real time in Infomobility is discussed in Section 4. The Smart Cities scenario is presented in Section 5. Conclusions and future tesbed extensions are discussed in Section 6.

2. The network infrastructure and QoS system

Although the main purpose of this work is to present measurements, it is important to describe the testbed setup and some related installation problems, as well as the QoS system and its main principles.

2.1 Testbed setup

The network plant is depicted in Fig. 1 and, from the left to the right, involves a Shelter for Internet distribution (A) which, due to property reasons, could not host the QoS Management Server (B), installed 1.2 km away (Link 1).

Such system runs on a Dual Xeon 2 GHz, 8Gb di Ram, 80Gb SAS Raid 5 server and comprehends a Radius authentication server, a PPPoE concentrator and the whole QoS management software.

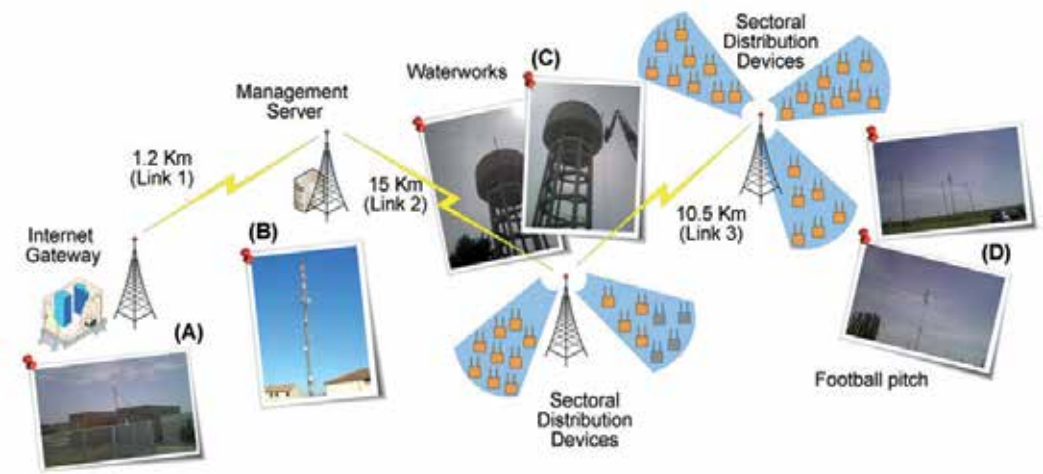


Fig. 1. the QoS equipment.

A HiperLAN link starts here and connects the management server to a waterworks area (C) 15 km away (Link 2). In such area, two sectoral distribution equipments were setup and serve 35 users approximately.

A further antenna allows to reach the last distribution area, situated near a football pitch (D) 10.5 km away (Link 3), and including three further sectoral distribution equipments for 45 users approximately. Such antenna had to be setup since some unfavourable features of the ground prevented a direct connection from being created.

The indirect link creates a bottleneck and forces the waterworks area to support some of the traffic surrounding the pitch.

A further penalty is that the same pylon which hosts the antenna in area (C) also carries some television and microwave aerials. In consequence, some devices not optimally shielded were initially blocked and even damaged and the available bit-rate is still being diminished.

In addition, some further installation problems were caused by the daily activation of the waterworks pump, which produced strong perturbations to the electric network, consequent blocks of many devices and even breakdowns. This problem was solved by means of an electronic filter.

All the above problems would have obviously prevented 80 users from being propely served, unless a bandwidth optimization method and traffic management were adopted.

2.2 The QoS architecture adopted

The QoS scheme is based on the dynamic assignment and redistribution of bandwidth on the basis of priority and users' profiles. The main parameters of each kind of profile are

summed up in Tab. 1. In particular, when QoS management is enabled in a profile (QoS flag), different priorities can be assigned to distinct protocols.

Parameter	Description
Name	Profile name
Description	Profile description
Upload bandwidth	Max upload bandwidth (kbit/s)
Upload guaranteed	Min upload bandwidth guaranteed (kbit/s)
Download bandwidth	Max download bandwidth (kbit/s)
Download guaranteed	Min download bandwidth guaranteed (kbit/s)
QoS	Flag for enabling QoS traffic management

Table 1. main fields of a connection profile.

Fig. 2 shows the graphic interface for the definition and management of each type of profile. The seven panels "Band 1",. .. "Band 7" on the right allow to assign priorities, 1 being the highest, 7 the lowest. In particular, a protocol can be associated to a specif priority by dragging and dropping its name in the chosen band panel.

The minimum bandwidth guaranteed (in percentage) and maximum available must also be setup for each sub-bandwidth. Once a profile has been defined, it can be assigned to many distinct users, so as to tailor service supply easily and quickly.

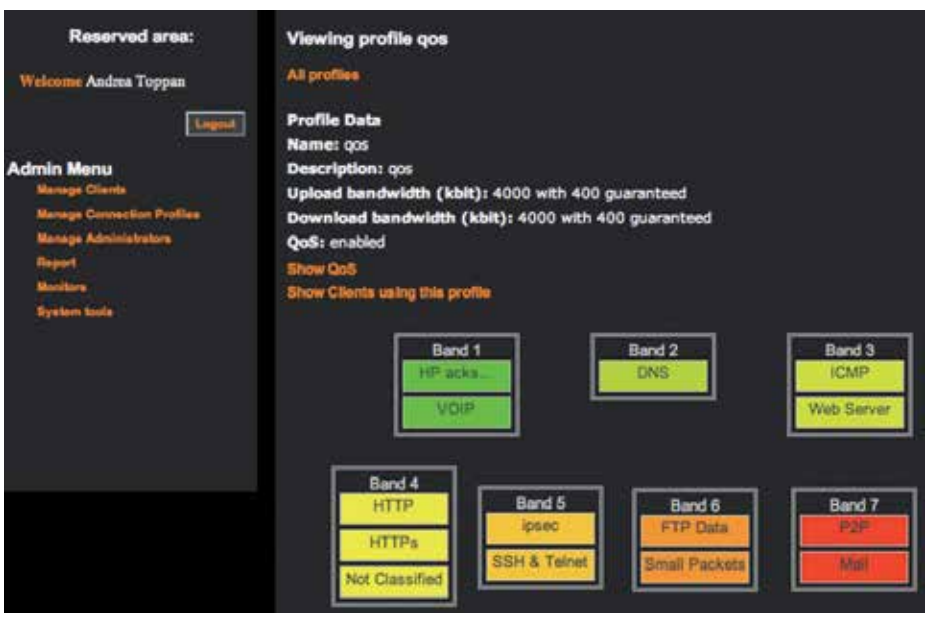


Fig. 2. priority assignment through drag&drop operations.

As far as dynamic QoS management is concerned, the basic idea is to limit both upload and download operations through the Egress policier (www.egress.com), so as to discard

all the packets whose speed exceeds a maximum value set. To this purpose, queuing algorithms are applied and bandwidth can consequently be tuned according to actual availability.

Some problems had to be solved along the way: shaping could only be applied to the outgoing traffic, already processed by the kernel, whereas both the uploading and downloading flows should normally be shaped by queuing methods on both the incoming and outcoming interfaces. In this case, though, many independent PPP interfaces were simultaneously active and each PPPoE was thus identified through a PP system interface numbered N (PPN, N = 0, 1,. ..).

In consequence, two further difficulties had to be faced:

1. too many iptables rules were generated and so was a further branch in the queuing structures;
2. the htp qdisc bandwidth sharing capabilities could not be fully exploited and no minimal bandwidth per PPP connection could be guaranteed.

As a matter of fact, each PPP having its own independent queuing, the traffic on the network interface was managed in an unpredictable way: no minimum bandwidth per connection could be even assigned and the unexploited bandwidth could not be dynamically and equally redistributed among all connections.

In order to handle the above situation, a qdisc (common to all connections) and subclasses for each kind of connection (with minimum and maximum bandwidth set) were defined, so as to redistribute unexploited resources among tunnels.

To this purpose, a hierarchical structure based on HTB (Hierarchical Token Bucket, http://luxik.cdi.cz/~devik/qos/htb) queuing was developed, whose nodes specify their own minimum and maximum bandwidth. In this way, the traffic of each tunnel is forwarded to the class it pertains to, so as to achieve the desired result.

Nevertheless, qdiscs can only manage the traffic of their own interface, so it was still impossible to identify a single connection by accessing the network interface of the PPPoE server. Each connection, in fact, is managed as a separate network interface.

An IMQ (Intermediate Queueing Device, www.linuximq.net) interface was thus adopted, which allows to manage qdiscs and the whole traffic: iptables are deviated to such interface and traffic can be shaped. Each single PPP interface must be assigned a connection identifier and sent to the IMQ, where connection classes were defined.

In this way, each connection can monitored, traffic can be classified on the basis of protocols and the most important flows are assigned the highest priority.

Another problem faced was that each packet could only be marked by means of an an identifier, so, theoretically speaking, the simultaneous identification of connections and protocols within a session was impossible.

Several tests demonstrated that the problem could be solved through the joint use of u32filter+MARK and CLASSIFY TARGET. This was done defining a further HTB class structure in the PPPN interfaces.

The bandwidth redistribution problem having been solved, attention could be focused on flow priority within each connection.

Fig. 3 shows how queuing algorithms were applied. A root node (qdisc) was created in the imq0 interface; class 1:1 was added in order to define the total bandwidth (100 Mbit/s in this case). Subclasses were then defined for the management of single connections, each specifying the minimum guaranteed (rate) and maximum at disposal (ceil).

Note that a higher QoS could be achieved if the SFQ (Stochastic Fair Queuing) were applied, so as to manage single flows through a Round Robin policy.

In order to manage priority of single flows, a hierarchic structure was created within each PPPN interface.

As in the former case, several subclasses were added to the root node (qdisc), each with a minimum and maximum bandwidth; the SFQ algorithm was applied afterwards. A u32filter list is defined in the root node of each interface, so as to drive single packets on the class they pertain to on the basis of their protocol. Packets were initially divided by the CLASSIFY TARGET tool according to the connection; a further division was then made by u32filter+MARK on the basis of protocols.

Besides, the identifiers range is 1:10000 and 1:65535 in hex, so the highest attention must be paid to each class handles. The following synatx was adopted:

```
iptables -t mangle -A POSTROUTING -j CLASSIFY --set-class x:y
```

In this way, filters could be avoided and everything is managed by iptables.

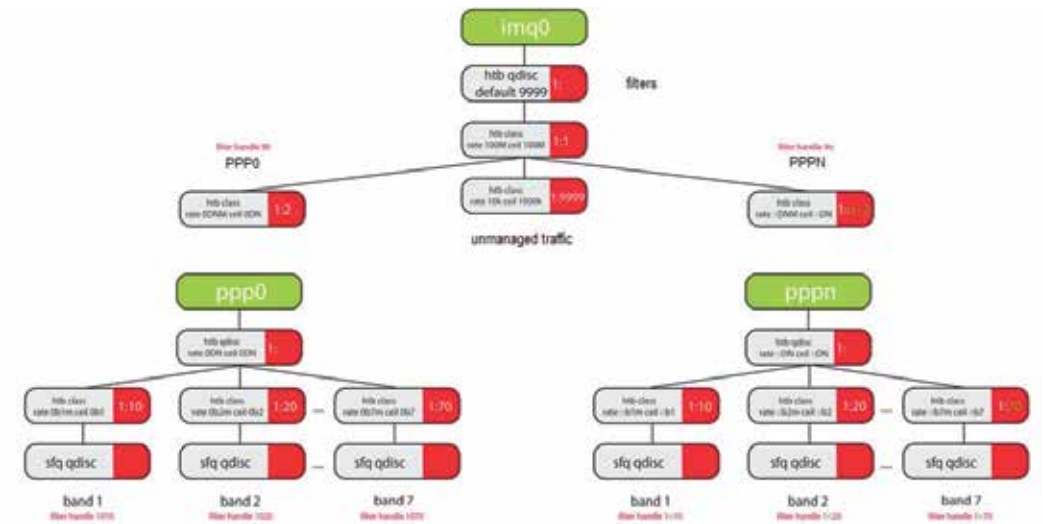


Fig. 3. hierarchical QoS management.

2.2.1 Statistics

The platform allows to visualize information about the infrastructure and its use and, in consequence, to make statistics about connected users and their traffic volume.

Fig. 4, for instance, refers to all the users' total connection time. For the sake of compactness, only a small excerpt was reported.

Diagrams are also available describing the system components, such as CPU load, network load and many others.

User_id	Connection time	Upload	Download	Online
[redacted]	161 days, 4 hours, 33 minutes, 10 seconds	78.69 GB	112.18 GB	ONLINE
[redacted]	12 days, 20 hours, 47 minutes, 6 seconds	19.51 MB	208.73 MB	OFFLINE
[redacted]	189 days, 11 hours, 59 minutes, 20 seconds	17.96 GB	33.32 GB	ONLINE
[redacted]	230 days, 23 hours, 25 seconds	0.81 GB	6.06 GB	ONLINE
[redacted]	105 days, 1 hours, 10 minutes, 13 seconds	130.13 MB	1.33 GB	ONLINE
[redacted]	187 days, 16 hours, 6 minutes, 24 seconds	51.65 GB	78.91 GB	ONLINE
[redacted]	104 days, 12 hours, 12 minutes, 25 seconds	3.81 GB	12.07 GB	ONLINE
[redacted]	191 days, 9 hours, 59 minutes, 53 seconds	2.31 GB	6.88 GB	ONLINE
[redacted]	123 days, 5 hours, 13 minutes, 23 seconds	18.01 GB	23.85 GB	ONLINE
[redacted]	13 days, 23 hours, 30 minutes, 18 seconds	31.34 MB	283.81 MB	ONLINE
[redacted]	250 days, 22 hours, 13 minutes, 34 seconds	3.00 GB	34.96 GB	ONLINE
[redacted]	96 days, 7 hours, 42 minutes, 36 seconds	9.28 GB	23.43 GB	ONLINE
[redacted]	200 days, 18 hours, 22 minutes, 15 seconds	12.37 GB	76.01 GB	ONLINE
[redacted]	95 days, 19 hours, 23 seconds	2.07 GB	8.91 GB	OFFLINE

Fig. 4. small excerpt from all the users' total connection time.

As far as each user is concerned (Fig. 5), the following data can be monitored: connections, data volumes exchange, diagrams about his or her traffic and, as indicated by law, packets logging.The authentication system adopted is Radius and traffic is encapsulated through the PPPoE protocol. In consequence, a PPP tunnel is active between each user and the server.

3. Measurements campaigns

As already anticipated, the main concern of this paper is to present a realistic testbed for QoS management; four measurements campaigns carried out are described in the following, whose scenario was described in Section 2.1 (Fig. 1).

The adopted dynamic priority-based method is twofold. On the one hand, users are assigned the shared bandwidth on the basis of their profiles: the higher the guaranteed bandwidth, the higher the shared bandwidth assigned. On the other hand, not only users but also services within a profile are prioritized, so each user is aware that his or her bandwidth is accordingly shared among his or her applications.

The QoS management is presented in an increased way: no QoS in the first campaign; QoS with neither minimum nor maximum set in the second; minimum and maximum bandwidth defined in the third; dynamic redistribution is also managed in the fourth.

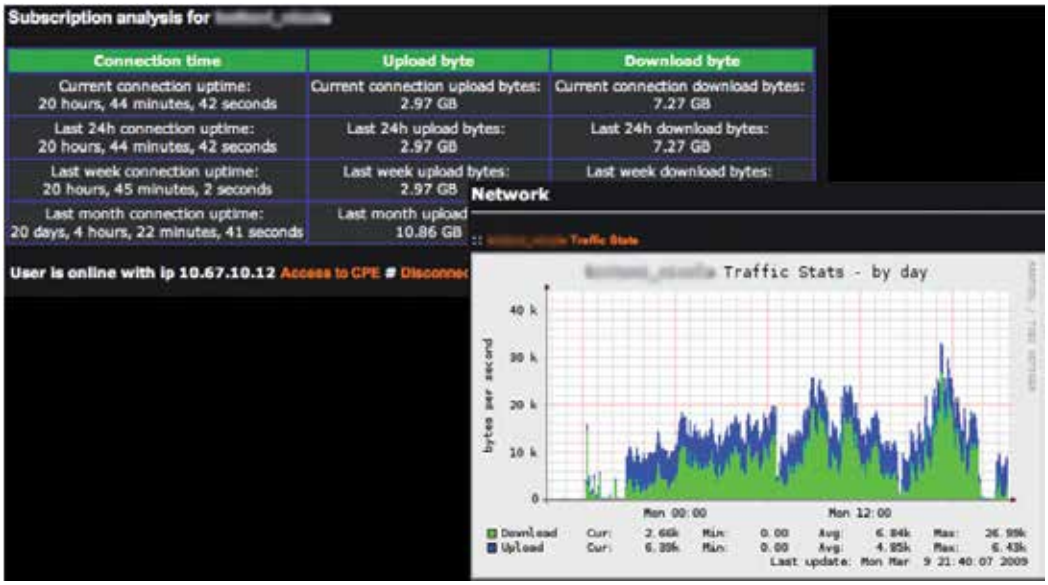


Fig. 5. statistics about a single user.

3.1 Throughput on single links (no QoS)

The first campaign represents throughput maximization in the single links of the whole infrastructure. Fig. 6 shows results in Link 1, from the Shelter to the Management Server.

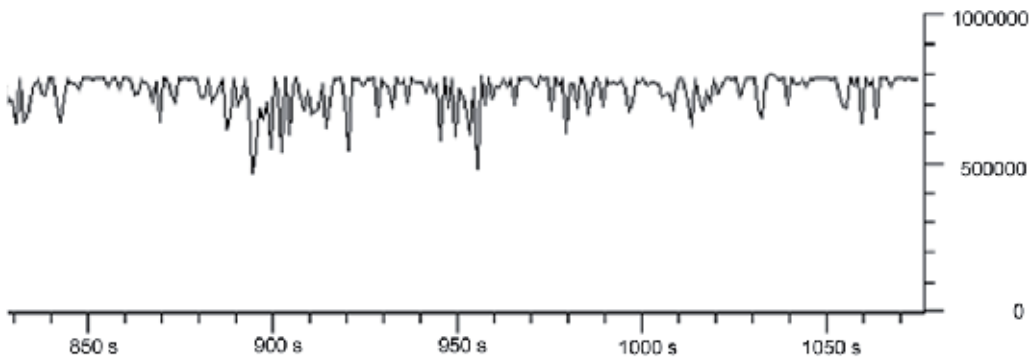


Fig. 6. throughput in the link from the Shelter to the Management Server (bit/s).

The above measurements were performed using the following tools: (1) Iperf (www.noc.ucf.edu/Tools/Iperf), which allows to send a TCP or UDP data streams and measure their throughput; (2) WireShark (www.wireshark.org), a network analyzer which allows to capture and diagram Iperf streams.

In this campaign, the Iperf server was installed on an Acer Travelmate with Linux Debian OS and located in the B node, so as to receive UDP connections. A Macbook Pro with Mac OSX Leopard was used as the client.

Nodes A, C, D hosted laptops for the connection to the Iperf server. In this way, throughput could be measured first in Link1, then Link2 and Link 3 and finally in Link2 + Link3.

The highest UDP throughput at disposal in Iperf connnections was set to 10 Mbit/s; results are reported in Tab. 2 and derive from a large number of measurements properly mediated.

Note that the throughput from the Management Server to the football pitch (Link 2 + Link 3) is almost 2Mbit/s lower than in single links 2 and 3: this derives from the the same device being charged of both signal reception and transmission.

Link	Description	Throughput
Link 1	From the Shelter to the Management Server	7.8 Mbit/s
Link 2	Management Server to Waterworks	6.6 Mbit/s
Link 3	Waterworks to Football pitch	6.9 Mbit/s
Link2 + Link 3	Management Server to Football pitch	5.2 Mbit/s

Table 2. results in single links.

3.2 QoS applied to multimedia TCP flows (no min/max bandwidth set)

The second campaign aimed at verifying the efficiency of the QoS management server. The Iperf was moved in the (C) node, so as to check Link 3.

An ethernet cable subsituted the wireless connection during a PPP connection. Delays and packet loss, in fact, are not particularly relevant in this kind of control, attention being mainly focused on flow management.

The following tools were adopted: (1) QoS Server; (2) Server-side Vlc for MMS over http video flow transmission (www.videolan.org/vlc); (3) Client-side Vlc for flow reception; (4) WireShark.

Fig. 7, diagrammed through WireShark, represents the scenario and the first measurements of this campaign. It refers to the following profile: no QoS applied, symmetric upload and dowmlod of 1Mbit/s, no bandwidth guaranteed. The client receives the first video (red line) until second 230, then the second video (green line) starts and the firts one is interrupted at second 280.

As expected, in the concurrent period (sec. 230 to 280) both TCP videos are blocked, the bandwidth being inadequate to support both of them.

Fig. 8 represents the second test and involves three videos on three distinct ports; in this case, a QoS profile was enabled which guarantees an increasing priority from the first to the third flow.

First starts the video on the lowest priority port (blue curve); the intermediate priority video starts 20 seconds later (green curve) and, in consequence, the first data flow declines. In the period between sec. 40 to 80 the maximum throughput was increased, so as to emphasize the effect of dark and still scenes in the second video. In this way, the total throughput is constant and bandwidth waste is kept under control.

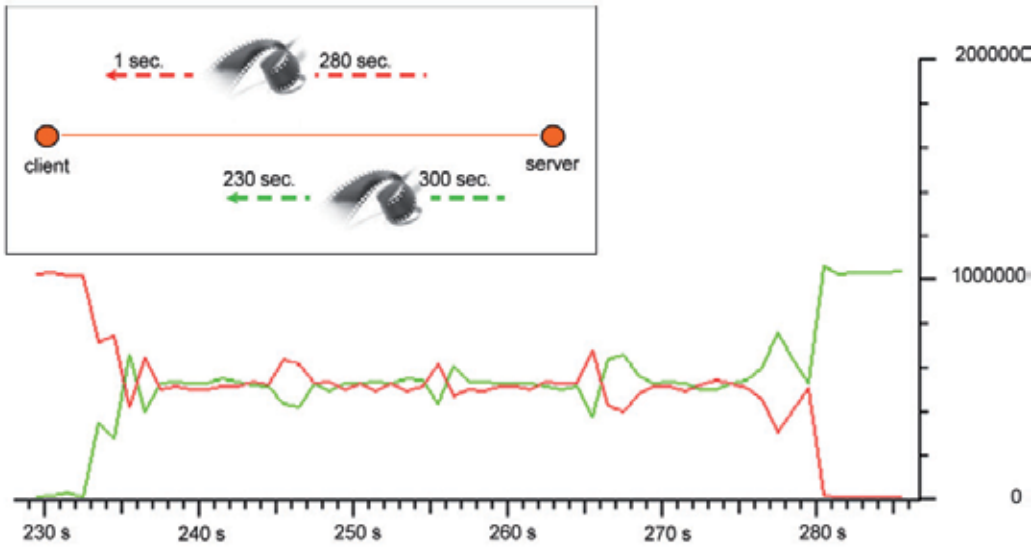


Fig. 7. two concurrent video flows without QoS (bit/s): scenario and results.

The same observations apply to the third and highest priority flow (red curve): the three videos share the bandwidth and, thanks to QoS, the lowest priority one is flattened, the medium is assigned less bandwidth and the highest has the best quality.

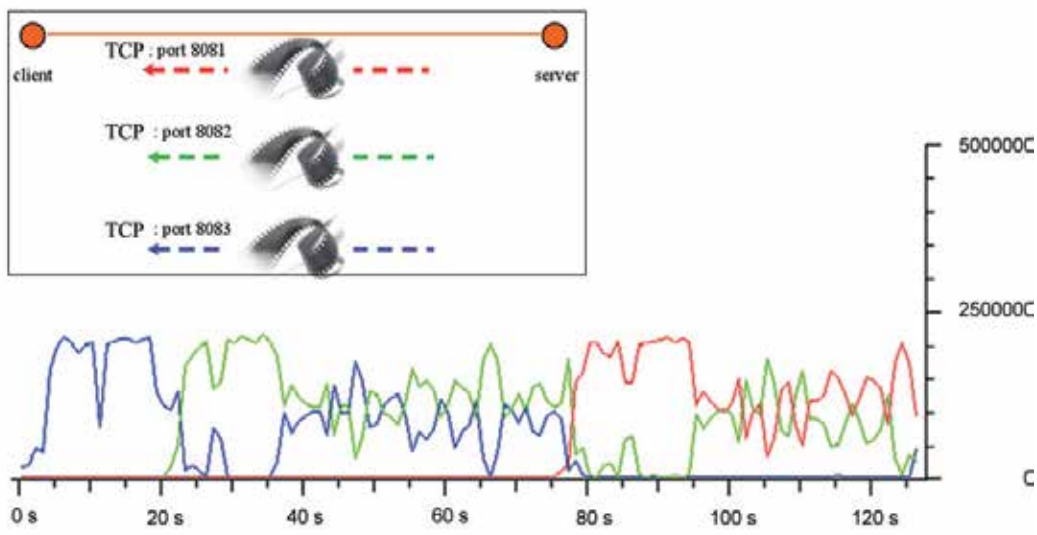


Fig. 8. three concurrent videos with QoS (bit/s).

3.3 Bandwidth control in QoS management

As the previous measurements showed, the QoS policy adopted helps to avoid bandwidth waste and guarantees a better service, especially for VoIP, IPTV support etc.

Nevertheless, this kind of priority management among traffic classes implies the almost complete cancellation of the less important services for the benefit of the most important ones.

In the initial QoS profile, an increasing priority was assigned to the three TCP flows on distinct ports. Fig. 9 shows the measures obtained. The less priority flow (blue curve) is strongly limited by the second one (green curve). They are both flattened when the highest priority flow starts (red curve).

In order to improve such results, each traffic class was then assigned a minimum and a maximum throughput (Tab. 3). The third campaign tries to demonstrate the effectiveness of such method and Fig. 10 reports the results.

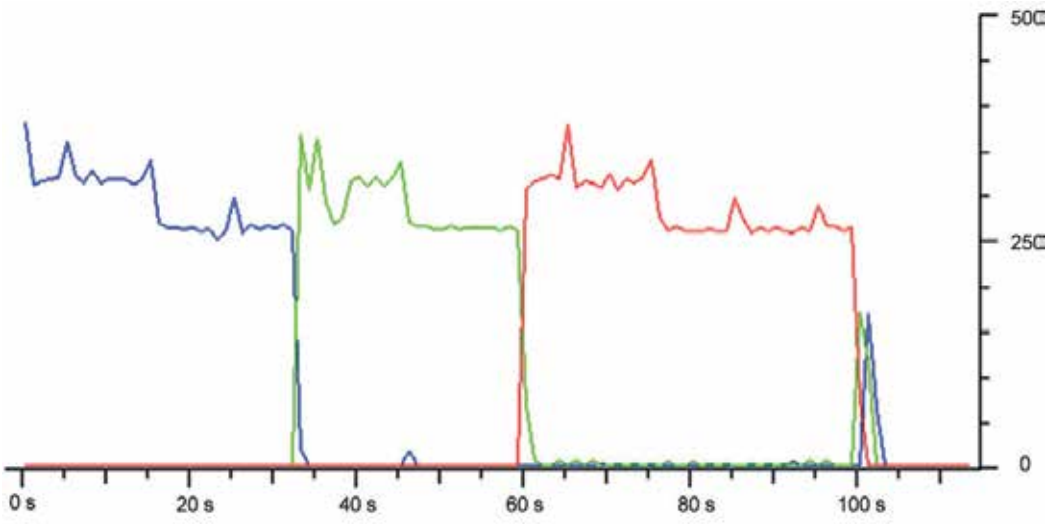


Fig. 9. TCP concurrent flows with QoS (bit/s).

Class	Priority level	Min. bandwidth % guaranteed	Max. bandwidth % at disposal
1	Highest priority	0%	100%
2	Intermediate	30%	30%
3	Lowest priority	20%	100%

Table 3. minimum throughput guaranteed and maximum available, as assigned to single flows.

Note that an upper bound having been imposed to the intermediate flow, the lowest is not totally flattened, but only slowed down.

The most priority flow starts at second 50 and, in consequence, the less priority traffic becomes slower, but not more than the 20% guaranteed. The same applies to the intermediate flow, for which a 30% at least is available. The highest priority flow, of course, can not reach the maximum speed.

As this measure shows, if guaranteed bandwidth percentages are properly managed, a high QoS can be obtained in an easy and immediate way.

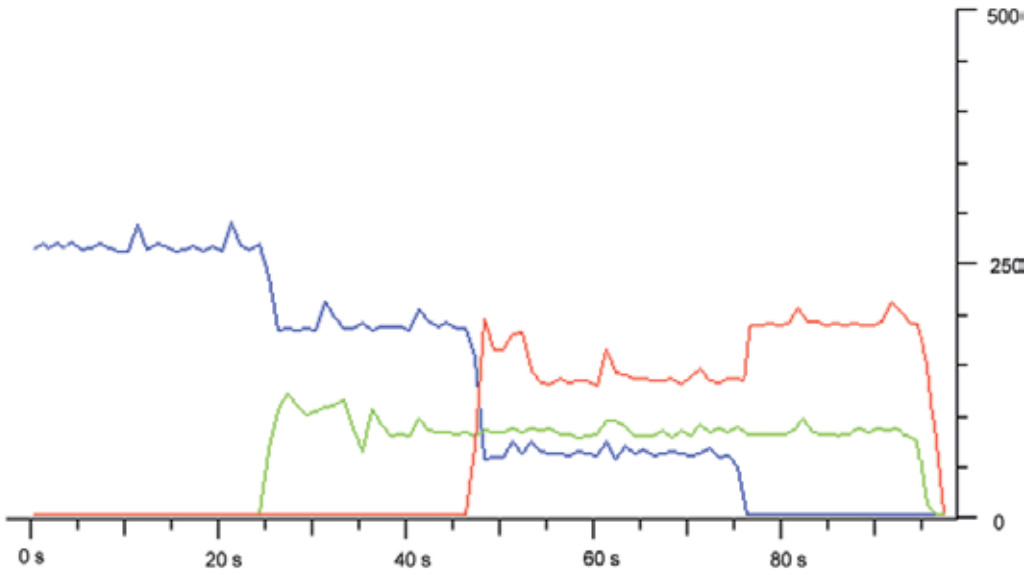


Fig. 10. TCP concurrent flows with bounded bandwidth QoS (bit/s).

3.4 Proportional reassignment of bandwidth

An important feature that must be handled in this kind of QoS management is bandwidth reassignment proportionally to each user's minimum guaranteed.

In this case, three PCs and the usual tools were adopted and two kinds of profiles were defined (Tab. 4).

Clients	Min. guaranteed	Max. at disposal
Clients 1 and 2	128 Kbit/s	3 Mbit/s
Client 3	600 Kbit/s	3 Mbit/s

Table 4. Minimum and maximum throughput for each client.

Clients were connected to the server through the PPPoE protocol; the maximum throughput between clients and server was set to 3Mbit/s, so as to simulate a set of wireless relays.

In this case, the upper bandwidth was to be shared among concurrent users and, in consequence, none was to reach the maximum.

Results are shown in Fig. 11: initially, the only connected client 1 (green curve) gets the whole bandwidth available according to his profile (3 Mbit/s).

After 100 seconds approximately, client 2 is also connected, throughput is assigned according to the minimum guaranteed and exceeding bandwitdh is reassigned according to such value.

The two clients share the same profile, so the bandwidth is equally divided and redistributed.

At second 150 the third client (red curve) connects to the system; the minimum throughput of his or her profile is four times the others', so clients 1 and 2 are limited accordingly.

When client 1 disconnects, bandwidth is distributed among clients 2 and 3 in a ratio of 1 to 4.

Finally, client 3 logs out and the whole bandwidth is at client's 2 disposal.

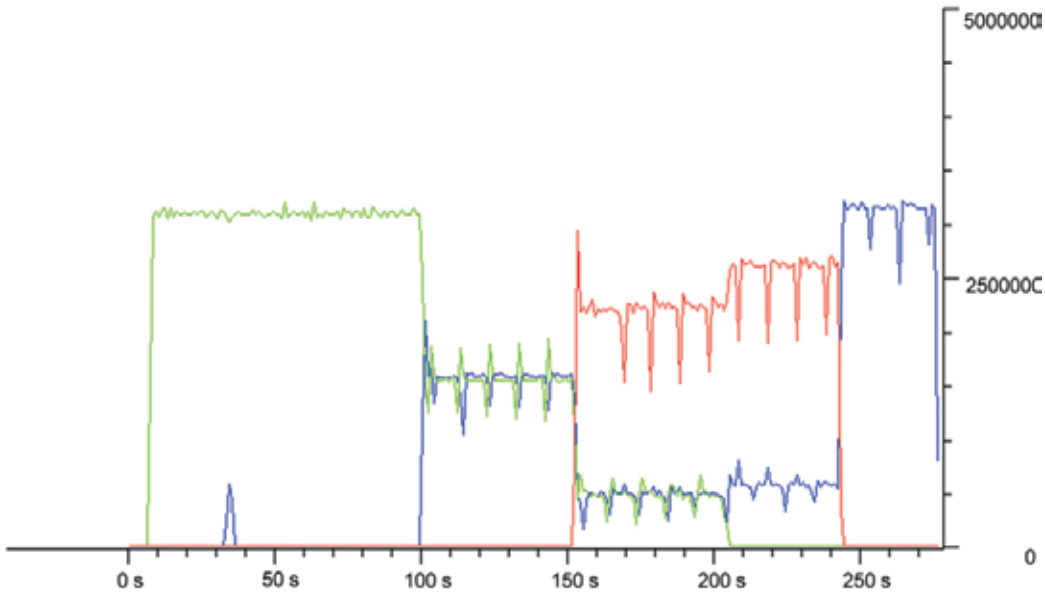


Fig. 11. bandwidth proportional reassignment among connected users.

4. The PEGASUS project: Real time support in infomobility services

One of the most important applications to which QoS techniques have been applied are Smart Navigation and the braoader field of Infomobility.

Transportation is one of the main fields where advanced technological systems can improve human life in a significant way: risks due to accidents, time wasted travelling and pollution could be highly reduced by applications for vehicle localization, behaviour prediction, etc.

These considerations are at the basis of the increasing interest that ITS are gaining in these years.

Furthermore, the latest study on global urbanization conducted by the Population Division of the Department of Economic and Social Affairs of the United Nations predicts that, in 2050, nearly 70% of the global population will be living in larger cities [23].

This immense aggregation of people will surely pose great challenges to the sustainability of modern lifestyle, and the problem of an efficient management of mobility stands out as one of the most relevant ones.

As a matter of fact, densely populated cities imply the concentration (from the country) and distribution (within the city) of massive amounts of people and resources [24].

In addition to the vast economic importance and consequences of such situation, urban and sub-urban mobility is a serious challenge also due to the circulation of large amounts of people and goods in a relatively small area. This poses hazards to life and health, especially for children, the elderly, and unfamiliar visitors, as well as to the environment.

Urban mobility, in fact, accounts for some 30% of energy consumption and 70% of transport pollution in Europe, and this problem is magnified by the increasing population concentration in large cities.

In such a scenario, the efficient management of traffic is a challenge that governments, industries and researchers are forced to face worldwide. Private travellers, commercial road users, and the public sector are continually searching for new and faster travel routes and methods.

Roads efficiency can be substantially improved by the deployment of ITS, which exploit ICT in order to provide traffic safety and efficiency.

ICT can be considered as the foundation for carrying out smart navigation, meant as the paradigm where mobile entities (vehicles and pedestrians) move wisely through a given environment, exploiting reliable and timely information about traffic conditions.

In this context, one of the most important applications is the support of real time, meant as the constant monitoring of traffic and road conditions, and the consequent possible update of the routes previously suggested. As a matter of fact, the best path in a given situation can vary when traffic conditions vary and updates should be notified to the user in real time. Nevertheless, up to now, no simple and marketable product was proposed for monitoring traffic and providing real time information to road users.

To this purpose, in the framework of the Italian project PEGASUS (http://pegasus.octotelematics.com/), WiLab aims at exploiting information transmitted from vehicles to a remote Control Center, so as to provide drivers in real time with updated information about actual traffic conditions. In this way, a new smart navigation service is supported. In particular, the objective is twofold:

- investigate the impact of smart navigation on the communication networks load;
- investigate the impact of real time updates on traffic management efficiency; as a matter of fact, vehicles equipped with smart navigators are constantly sent information about actual roads conditions;

In Fig. 12, the smart navigation scenario considered and developed at WiLab is shown: vehicles are equipped with on-board units (OBUs), which periodically transmit their speed and position (known through the GPS integrated on board) to a Control Center. Such data are transferred through the General Packet Radio Service (GPRS) network.

The fleet equipped with OBUs is addressed as *floating car data (FCD)*. In March 2010, the Italian FCD to which the PEGASUS project refers, reached over 1.000.000 equipped vehicles (OctoTelematics, 2010); this number is to increase quickly (note that the number of public

and private vehicles in Italy was 34 million back to 2003 [25], hence the FCD is a not negligible percentage of the overall private vehicles number).

All such data are processed and exploited for the real time dynamic navigation of vehicles (hereafter Dynamic Route Guidance, DRG); the same information can also be forwarded to public or private institutions for traffic management, etc.

In the near future, almost all vehicles will be able to send real time information, and the majority of drivers will take profit of data properly processed and of applications beyond traffic management, such as safety and entertainment.

In this scenario, telecommunications systems will be required to transmit information quickly and reliably, both among vehicles and between vehicles and remote control centers. Which technologies are to be chosen, how priorities must be managed, which capacity is required, are still open issues.

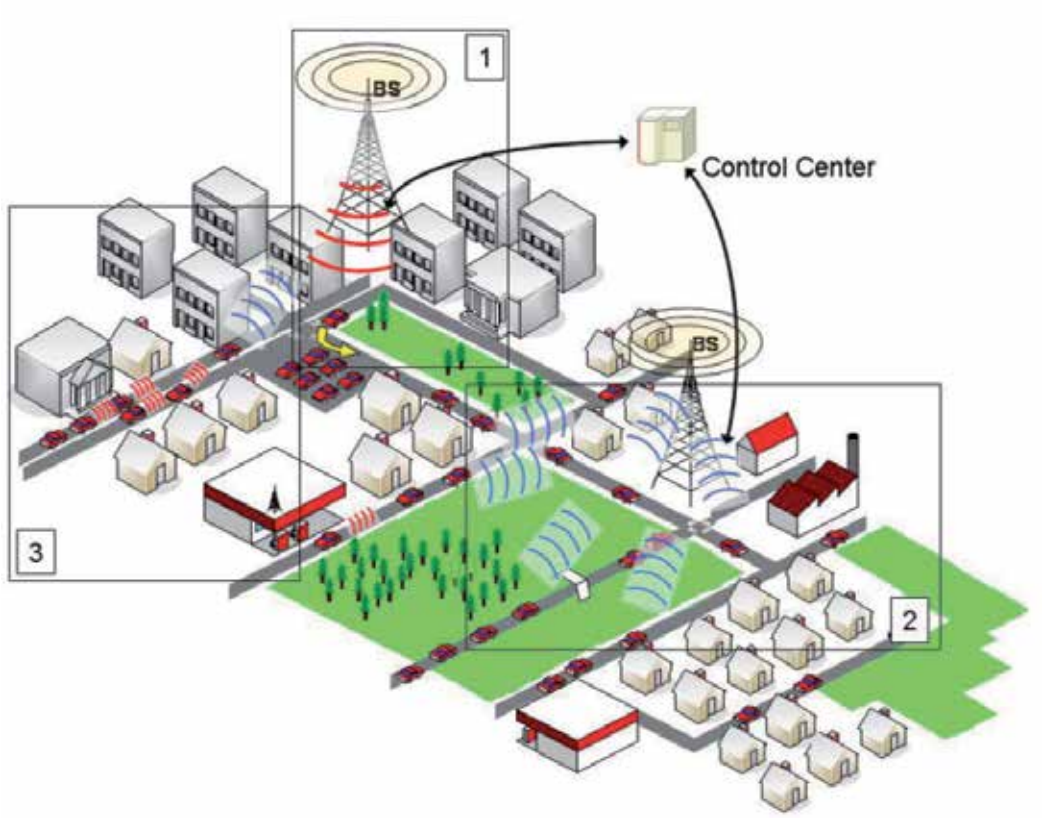


Fig. 12. Smart navigation scenario.

The mobile network is, at present, the only one adopted for vehicles-Control Center communication; nevertheless, the quantity and size of information is to increase.

Urban networks, thus, can turn out to be a precious support for existing infrastructures, especially if properly managed through effective QoS techniques.

5. The smart cities scenario

The QoS testbed is currently being applied to Smart Cities scenarios, a class of applications which are gaining an increasing attention.

As described by William J. Mitchell (MIT, Smart Cities Group, http://cities.media.mit.edu/): "Our cities are fast transforming into artificial ecosystems of interconnected, interdependent intelligent digital organisms. Emerging applications in the ICT field are poised to reshape our urban environments".

In this context, wireless architectures and QoS infrastructures become nodes of a TLC network, which collects information from the surrounding areas and consequently supplies citizens with advanced infomobility services.

A very diffused and challenging problem to face is where hardware can be installed, since many areas and city centers are protected and regulated by severe rules by the Ministry of Culture and Heritage.

A possible solution is to place the whole hardware within preexistent structures, such as electricity posts, properly adapted in order to host the required technology.

A testbed based on this approach adopts "Intelligent Posts" hosting both hardware and software (Fig. 13). Such posts were installed by the WiLab group, in cooperation with Fondazione Almamater and Ghisamestieri within Villa Gandolfi Pallavicini (Bologna, Italy).

Fig. 13. Posts by Ghisamestieri for Smart Cities.

In this case, the effective management of QoS is tested in order to supply citizens with several services, such as video surveillance (both wired and wireless data transmission involved), integrated image analysis, Internet connection within urban environments, RFID services for tourists, emergency calls, radiodiffusion, fire prevention, parking management, localization, diagnostics and control by television. In addition, sensor network data collection, traffic information, access control, mobile payments, vehicle tracking, user-generated contents, energy management, etc.

Through QoS management, such services will be configured dynamically on the basis of bandwidth availability. According to the throughput actually available in a specific temporal slot and thanks to a constant monitoring of radio resources on each route, both services to be offered to the user and applications to be kept active can be chosen.

More specifically, the testbed is addressed to transport improvement and traffic reduction through smart navigators.

6. Conclusions and future QoS testbed extensions

Coming back to the tests in Sections 2, 3, PPP tunnels between the server and users can be temporarily closed, when packet transimission is slowed down by interference or machinery stops.

The first problem is that, in case the PPP LCP surveys trouble situations, the channel is closed and the client disconnected. An authomatic procedure is in charge of reconnection, but a time waste in PPP tunnel setup as well as abrupt disconnections are bound to take place.

A second problem derives from traffic limitation and control being handled by a single QoS server: in this case, data are properly limited only after they have crossed one or two links. In other words, in case an authenticated user sends an UDP data flow larger than his or her maximum upload bandwidth, such flow will be diminished only after reaching the QoS server. Meanwhile, the available bandwidth will be unproperly occupied by such flow.

On the basis of such considerations, the testbed will be extended according to two different scenarios of distributed QoS architectures [26-28]. The first one is depicted in Fig. 14 and aims at avoiding tunnel closure in case of interference and packet loss.

In case many relays occur between the client and PPPoE concentrator, packet loss can increase; the idea, thus, is to shorten the tunnel, so as to integrate the PPPoE concentrator and the transmitter. In this way, all PPP features could be maintained and its limitations diminished. The tunnel, in fact, would be established between the client's CPE and the nearest transmitter and communication between the transmitter and the main server could be based on TCP/IP.

Furthermore, if interferences between transmitter and main server would take place, packets could be relayed without PPP tunnel drops.

A disadvantage could concern uncoded communication between the main server and pylons. Possible solutions could be the activation of encrypted systems or a PPP tunnel to the main server. In this case, the user would not even perceive any link failure.

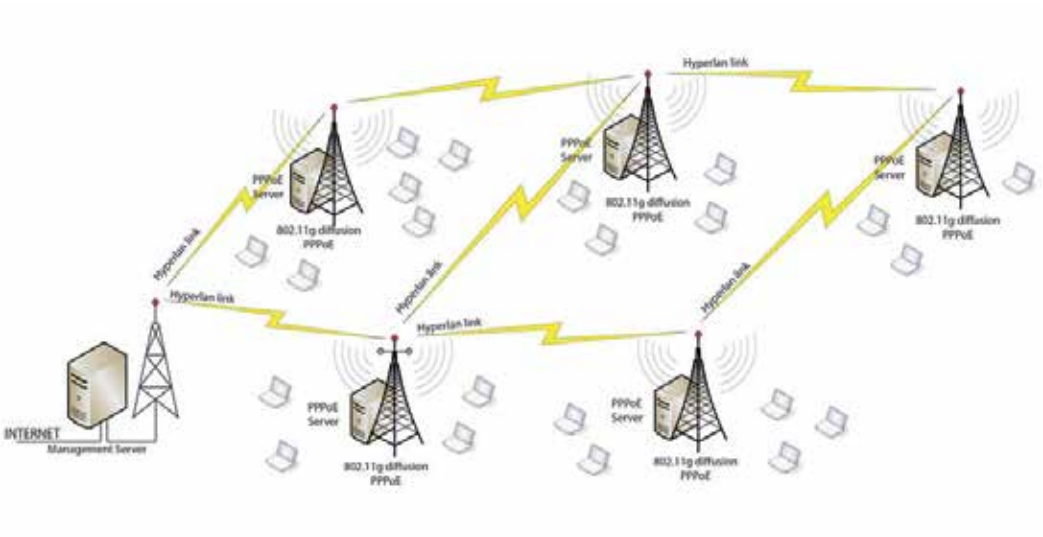


Fig. 14. first extended testbed scenario.

The second scenario (Fig. 15) aims at solving the second problem arisen: the idea is to apply the first control on users' bandwidth at the pylon.

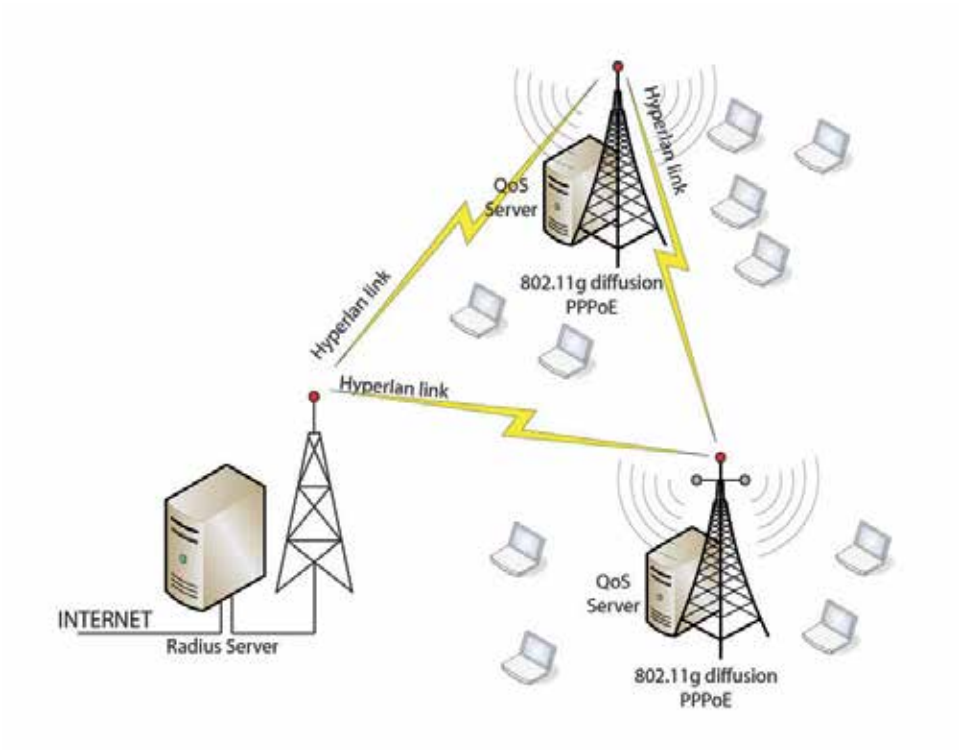


Fig. 15. second extended testbed scenario.

As for the PPPoE concentrator, the QoS manager itself could be integrated in the transmitter. On the one hand, this kind of control logic decentralization would solve the problem of link saturation in case of heavy UDP uploads. On the other hand, the server would be spared from an exceeding traffic in case of network expansion.

Two further difficulties arise: firstly, connection plans are not anymore managed by a single server in a centralized and transparent way. In consequence, a new communication protocol is required for the authomatic configuration of devices on the pylon when the main server configuration changes.

In addition, logic decentralization can cause more frequent failures of important components. If a failure occurs of QoS or PPPoE components, thus, an infrastructure is needed that prevents connections from being denied.

A switch, for instance, could be used to disconnect out of order devices and the main server would be in charge of guaranteeing connectivity until the problem is solved.

At present, the idea is to avoid a complete implementation of the above scenarios, using simulation tools instead. In particular (Fig. 16), the QoS impact could be evaluated through the joint use of the actually developed parts and simulators.

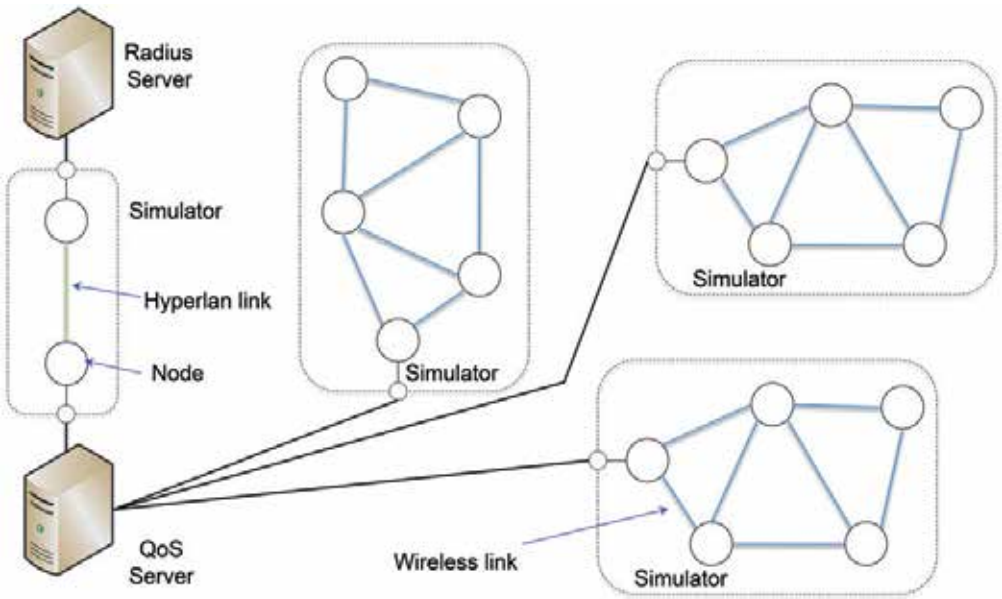


Fig. 16. A schema for the simulation of QoS server impact.

7. Acknowledgment

More than an acknowledgment, a dedication: To the little Gabriele Toppan, the son of Paolo and the nephew of Andrea, go our very best wishes to grow up strong, responsible and enthusiastic about life and its numerous miracles.

8. References

[1] Gringeri, S.; Shuaib, K.; Egorov, R. et al.; Traffic shaping, bandwidth allocation, and quality assessment for MPEG video distribution over broadband networks, Network, IEEE, 12 , 6, pp. 94 – 107, 1998, doi: 10.1109/65.752648

[2] Frank Yong Li; Stol, N.; QoS provisioning using traffic shaping and policing in 3rd-generation wireless networks, in Proc. of IEEE Wireless Communications and Networking Conference, 2002, 1, 139 – 143, 2002, doi: 10.1109/WCNC.2002.993478

[3] Yongdong Wang; Jurczyk, M.; Impact of traffic shaping in ATM networks on video quality, in Proc. of International Workshops on Parallel Processing, 1, 485 – 492, 2000, doi: 10.1109/ICPPW.2000.869154

[4] Gozalvez, D.; Monserrat, J.F.; Calabuig, D., et al; Policy-based channel access mechanism selection for QoS provision in IEEE 802.11e, Vehicular Technology Magazine, IEEE, 2, 3, 29-34, 2007, doi: 10.1109/MVT.2008.915326

[5] Flegkas, P.; Trimintzios, P.; Pavlou, G.; A policy-based quality of service management system for IP DiffServ networks, Network, IEEE, 16, 2, 50-56, 2002, doi: 10.1109/65.993223

[6] Fangming Zhao; Lingge Jiang; Chen He; Policy-based radio resource allocation for wireless mobile networks, in Proc. of IEEE International Conference on Neural Networks and Signal Processing, 476-481, 2008, doi: 10.1109/ICNNSP.2008.4590396

[7] Conchon, E.; Pérennou, T.; Garcia, J. Et al.; W-NINE: A Two-Stage Emulation Platform for Mobile and Wireless Systems, EURASIP Journal on Wireless Communications and Networking, 2010, Article ID 149075

[8] Heithecker, S.; do Carmo Lucas, A.; Ernst, R.; A High-End Real time Digital Film Processing Reconfigurable Platform, EURASIP Journal on Embedded Systems, 2007, Article ID 85318

[9] Chang Wook Ahn; Ramakrishna, R.S.; QoS provisioning dynamic connection-admission control for multimedia wireless networks using a Hopfield neural network, Vehicular Technology, IEEE Transactions on, 53, 1, 106-117, 2004, doi:10.1109/TVT.2003.822000

[10] Xiang Chen; Bin Li; Yuguang Fang; A dynamic multiple-threshold bandwidth reservation (DMTBR) scheme for QoS provisioning in multimedia wireless networks, Wireless Communications, IEEE Transactions on, 4, 2, 583-592, 2005, doi: 10.1109/TWC.2004.843053

[11] Huan Chen; Kumar, S.; Kuo, C.J.; Dynamic call admission control scheme for QoS priority handoff in multimedia cellular systems, in Proc. of IEEE Wireless Communications and Networking Conference, 114-118, 2002, doi: 10.1109/WCNC.2002.993474

[12] Yaya Wei; Chuang Lin; Fengyuan Ren et al; Dynamic priority handoff scheme in differentiated QoS wireless multimedia networks, in Proc. of Eighth IEEE International Symposium on Computers and Communication, 131-136, 2003, doi: 10.1109/ISCC.2003.1214112

[13] Xiaorong Li; Chuah, E.; Jo Yew Tham; Kwong Huang Goh; An optimal smooth QoS adaptation strategy for QoS differentiated scalable media streaming, in Proc. of IEEE International Conference on Multimedia and Expo, 429-432, 2008, doi: 10.1109/ICME.2008.4607463

[14] Huang, J.-H.; Li-Chun Wang; Chung-Ju Chang; Capacity and QoS for a Scalable Ring-Based Wireless Mesh Network, IEEE Journal on Selected Areas in Communications, 24, 11, 2070-2080, 2006, doi: 10.1109/JSAC.2006.881622

[15] Bai, B; Chen, W.; Cao, Z. et al; Uplink Cross-Layer Scheduling with Differential QoS Requirements in OFDMA Systems, EURASIP Journal on Wireless Communications and Networking, 2010, Article ID 168357

[16] Montazeri, S.; Fathy, M.; Berangi, R.; An Adaptive Fair-Distributed Scheduling Algorithm to Guarantee QoS for Both VBR and CBR Video Traffics on IEEE 802.11e WLANs, EURASIP Journal on Advances in Signal Processing, 2008, Article ID 264790

[17] Almeida, M.; Sarrô, R.; Barraca, J.P. et al; Experimental Evaluation of the Usage of Ad Hoc Networks as Stubs for Multiservice Networks, EURASIP Journal on Wireless Communications and Networking, 2007, Article ID 62967

[18] Ganesh Babu, T.V.J.; Le-Ngoc, T.; Hayes, J.F.; Performance of a priority-based dynamic capacity allocation scheme for wireless ATM systems, IEEE Journal on Selected Areas in Communications, 19, 2, 355-369, 2001, doi: 10.1109/49.914513

[19] Naser, H.; Mouftah, H.T.; A class-of-service oriented packet scheduling (COPS) algorithm for EPON-based access networks, in Proc. of 7th Int. Conference on Transparent Optical Networks, 232-236, 2005, doi: 10.1109/ICTON.2005.1505793

[20] Song, S.; Manikopoulos, C.N.; A Priority-based Feedback Flow Control System for Bandwidth Control, in Proc. of 40th Annual Conference on Information Sciences and Systems, 1645-1652, 2006, doi: 10.1109/CISS.2006.286399

[21] Zhang, F.; Verma, P.K.; Cheng, S.; Pricing, resource allocation and quality of service in multi-class networks with competitive market model, Communications, IET, 5, 1, 51-60, doi: 10.1049/iet-com.2009.0694

[22] Kamosny, D.; Novotyny, V.; Balik, M.; Bandwidth Redistribution Algorithm for Single Source Multicast Networking, in Proc. of Int. Conference on Systems and Int. Conference on Mobile Communications and Learning Technologies, 147-156, 2006, doi: 10.1109/ICNICONSMCL.2006.62

[23] UN, World urbanization prospects: The 2007 revision population database, 2008, http://esa.un.org/unup/

[24] EU, Eu mobility and transport, 2010, http://ec.europa.eu/transport/publications/statistics/

[25] ecoage, Independent ecology portal, 2003, www.ecoage.net

[26] Won-Kyu Hong, D., Choong Seon Hong, C.; A QoS management framework for distributed multimedia systems, Int. J. Network Mgmt, 13, 115-127, 2003, doi: 10.1002/nem.465

[27] Jing Li; Yongwang Zhao; Min Liu et al; An adaptive heuristic approach for distributed QoS-based service composition, in Proc. of IEEE Symposium on Computers and Communications (ISCC), 687-694, 2010, doi: 10.1109/ISCC.2010.5546721

[28] Pattara-Atikom, W.; Krishnamurthy, P.; Banerjee, S.; Comparison of distributed fair QoS mechanisms in wireless LANs, in Proc. of IEEE Global Telecommunications Conference GLOBECOM '03. 553-557, 2003, doi: 10.1109/GLOCOM.2003.1258298

5

End to End Quality of Service in UMTS Systems

Wei Zhuang
China Telecom Co. Ltd. (Shanghai)
P.R.China

1. Introduction

About ten years ago, WCDMA [1] based the third generation mobile systems started to be deployed worldwide. Besides to support basic mobile data services such as file transfer and internet surfer, etc., UMTS has one of the most significant archievements which can support a richer variety of services with QoS guarantee, such as video, VOIP, etc. Quality of Service (QoS) is defined as "the collective effect of service" performance, which determines the degree of satisfaction of a user of the service in the ITU-T recommendation E.800. At a technical level, QoS can be characterized by service availability, delay, jitter, throughput, packet loss rate (Nortel White Paper, 2002).

3GPP has put many efforts to define and standardize a QoS framework for data services, specially IP-based services. The standardization of a UMTS QoS model started in 1999. the development was based on the following key principles: operation and QoS provisioning needed to be possible in the wireless environment, usage of the Internet QoS mechanisms, applications and interoperability. This chapter is aimed to provide an overview of the UMTS end-to-end QoS architecture, describe how the QoS requirements to be realized from top layer to wireless links.

2. WCDMA QoS architecture

QoS standardization in UMTS PS domain enables UMTS to provide data service with end-to-end QoS guarantees. 3GPP proposed a layered architecture for supporting end-to-end QoS. It includes the following key elements (Sudhir Dixit et al., 2001):

- Mapping of end-to-end services provided by the UE, UTRAN, Core Network (CN), and external IP networks;
- Traffic classes and associated QoS parameters;
- Location of QoS functions;
- QoS negotiation;
- Multiplexing of flows onto network resources;
- An end-to-end data delivery model.

[1] Wideband Code Division Multiple Access W-CDMA - the radio technology of UMTS - is a part of the ITU IMT-2000 family of 3G Standards.

The layered UMTS QoS architecture is shown in Figure 1. The UMTS network can provide end-to-end QoS services from a Terminal Equipment (TE) to another TE. A network bearer service describes how to realize a certain network QoS. It is defined by the control signaling, user traffic transport and QoS management functionality, which enabling the provision of a contracted QoS (Sudhir Dixit et al., 2001; 3GPP23107, 2011). As the end-to-end service is conveyed over several networks, the end-to-end bearer service consists of different network bearer services. The end-to-end bearer service can be decomposed into TE/MT local bearer service, the UMTS bearer service and the external bearer service.

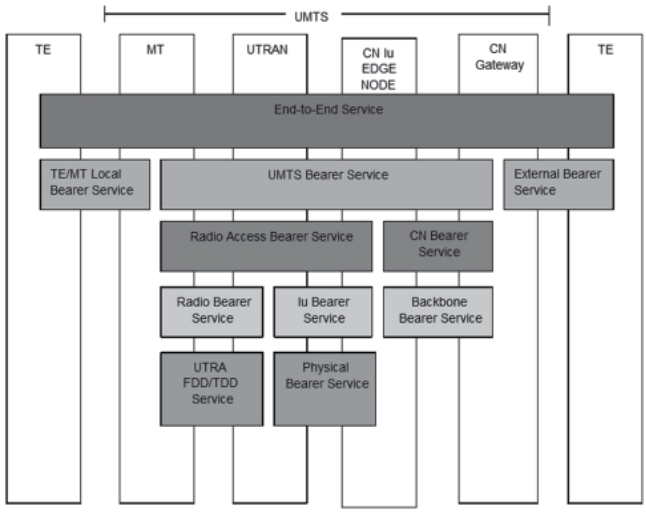


Fig. 1. UMTS QoS architecture

The TE/MT local bearer service provides communication between the TE and MT parts. MT (Mobil Terminal) provides connection to the UTRAN with basic functions, such as radio attachment to 3G network, authenticating the CS/PS domain, mobility management, etc. TE support call control, authenticating the IMS subscription, etc.

The external bearer service deals with the interoperability and interworking aspects with external IP bearer, and provides the appropriate functionality to support it. It is logical located in the GGSN, which is the gateway of UMTS to external network (Sotiris et al., 2002).

UMTS bearer service provides service by using the radio access bearer service (RAB) and the core network bearer service. The detail is given in the following.

2.1 UMTS bear service

The UMTS QoS is provided by the UMTS bearer service. It includes the radio access bearer service and the core network bearer service. They reflect the optimized way to realize the UMTS Bearer Service over the respective cellular network topology taking into account aspects such as mobility and mobile subscriber profiles.

The Radio Access Bearer Service provides confidential transport of signalling and user data between MT and CN Iu Edge Node with the QoS adequate to the negotiated UMTS Bearer Service or with the default QoS for signalling. This service is based on the characteristics of the radio interface and is maintained for a moving MT.

The Core Network Bearer Service of the UMTS core network connects the UMTS CN Iu Edge Node with the CN Gateway to the external network. The role of this service is to efficiently control and utilise the backbone network in order to provide the contracted UMTS bearer service. The UMTS packet core network shall support different backbone bearer services for variety of QoS. And the UMTS bearer service is realized by a GPRS service in the PS domain or a speech/data service in the CS domain.

2.1.1 The radio bearer service and Iu bearer service

The Radio Access Bearer Service is realised by a Radio Bearer Service and an Iu-Bearer Service. The role of the Radio Bearer Service is to cover all the aspects of the radio interface transport. This bearer service uses the UTRA FDD/TDD, which is not elaborated further in this chapter.

The Iu-Bearer Service together with the Physical Bearer Service provides the transport between UTRAN and CN. Iu bearer services for packet traffic shall provide different bearer services for variety of QoS.

2.1.2 The backbone network service

The Core Network Bearer Service uses a generic Backbone Network Service. The Backbone Network Service covers the layer 1/Layer2 functionality and is selected according to operator's choice in order to fulfill the QoS requirements of the Core Network Bearer Service. The Backbone Network Service is not specific to UMTS but may reuse an existing standard.

2.2 QoS requirement

2.2.1 UMTS QoS classes

The layered UMTS QoS architecture requires the definition of QoS attributes for each bearer service. When defining the UMTS QoS classes, the restrictions and limitations of the radio interface have to be taken into account. The QoS mechanism should be simpler than that in wired network due to different error characteristics of the air interface. Table 1 illustrates the QoS classes defined by 3GPP.

The main distinguishing factor between these QoS classes is how delay sensitive the traffic is.

Conversational class

The transfer time of real time conversation scheme shall be low because of the conversational nature of the scheme and at the same time that the time relation (variation) between information entities of the stream shall be preserved in the same way as for real time streams. The maximum transfer delay is given by the human perception of video and audio conversation. Therefore the limit for acceptable transfer delay is very strict, as failure to provide low enough transfer delay will result in unacceptable lack of quality. The transfer delay requirement is therefore both significantly lower and more stringent than the round trip delay of the interactive traffic case. The fundamental characteristic for QoS is to preserve time relation (variation) between information entities of stream and conversational pattern (stringent and low delay) (3GPP23107, 2011).

The most well known use of this scheme is telephony speech (e.g. GSM). But with Internet and multimedia a number of new applications will require this scheme, for example voice over IP and video conferencing tools. Real time conversation is always performed between

Traffic class	Conversational class conversational RT	Streaming class streaming RT	Interactive class Interactive best effort	Background Background best effort
Fundamental characteristics	- Preserve time relation (variation) between information entities of the stream Conversational pattern (stringent and low delay)	- Preserve time relation (variation) between information entities of the stream	- Request response pattern - Preserve payload content	- Destination is not expecting the data within a certain time - Preserve payload content
Example of the application	voice	streaming video	Web browsing	background download of emails

Table 1. UMTS QoS classes

peers (or groups) of live (human) end-users. This is the only scheme where the required characteristics are strictly given by human perception.

Streaming class

Streaming class is characterised by that the time relations (variation) between information entities (i.e. samples, packets) within a flow shall be preserved, although it does not have any requirement on low transfer delay (3GPP23107, 2011). This scheme is one of the newcomers in data communication, raising a number of new requirements in both telecommunication and data communication systems. It is a one-way transport. A user can use this class to watch (listen to) real time video (audio).

The delay variation of the end-to-end flow shall be limited, to preserve the time relation (variation) between information entities of the stream. But as the stream normally is time aligned at the receiving end (in the user equipment), the highest acceptable delay variation over the transmission media is given by the capability of the time alignment function of the application. Acceptable delay variation is thus much greater than the delay variation given by the limits of human perception. The fundamental characteristics for streaming class QoS is to preserve time relation (variation) between information entities of the stream.

Interactive class

Interactive traffic is the other classical data communication scheme that on an overall level is characterized by the request response pattern of the end-user. At the message destination there is an entity expecting the message (response) within a certain time. Round trip delay time is therefore one of the key attributes. Another characteristic is that the content of the packets shall be transparently transferred (with low bit error rate). The fundamental characteristics for interactive class QoS is to request response pattern, preserve payload content.

This class is applied when an end-user (either a machine or a human) is using on line requesting data from remote equipment (e.g. a server). Examples of human interaction

with the remote equipment are: web browsing, data base retrieval, server access. Examples of machines interaction with remote equipment are: polling for measurement records and automatic data base enquiries (tele-machines).

Background class

Background traffic is one of the classical data communication schemes that on an overall level is characterised by that the destination is not expecting the data within a certain time. The scheme is thus more or less delivery time insensitive. Another characteristic is that the content of the packets shall be transparently transferred (with low bit error rate). The fundamental characteristics for background class QoS are a) destination is not expecting the data within a certain time; b) preserve payload content.

When the end-user, that typically is a computer, sends and receives data-files in the background, this scheme applies. Examples are background delivery of E-mails, SMS, download of databases and reception of measurement records.

2.2.2 UMTS bearer service attributes

UMTS bearer service attributes describe the service provided by the UMTS network to the user of the UMTS bearer service. A set of QoS attributes (QoS profile) specifies this service. At UMTS bearer service establishment or modification different QoS profiles have to be taken into account.

Traffic class ('conversational', 'streaming', 'interactive', 'background')

Traffic class is a type of application for which the UMTS bearer service is optimized. By including the traffic class itself as an attribute, UMTS can make assumptions about the traffic source and optimise the transport for that traffic type.

Maximum bit-rate (kbps)

Maximum bit-rate (kbps) is the maximum number of bits delivered to UMTS at a SAP (Service Access Point) within a period of time, divided by the duration of the period. The traffic is conformant with Maximum bit-rate as long as it follows a token bucket algorithm where token rate equals Maximum bit-rate and bucket size equals Maximum SDU size.

The algorithm is well known as "Token Bucket Algorithm" which has been described in IETF. It is a reference algorithm for the conformance definition of bitrate. This may be used for traffic contract between UMTS bearers and external network/user equipment. In the algorithm, "tokens" represents the allowed data volume, for example in byte. "Tokens" are given at a constant "token rate" by a traffic contract, are stored temporarily in a "token bucket", and are consumed by accepting the packet. This algorithm uses the following two parameters (r and b) for the traffic contract and one variable (TBC) for the internal usage.

- r: token rate, (corresponds to the monitored Maximum bitrate/Guaranteed bitrate).
- b: bucket size, (the upper bound of TBC, corresponds to bounded burst size).
- TBC (Token bucket counter): the number of given/remained tokens at any time.

According to a token bucket, conformance can be defined as: "Data is conformant if the amount of data submitted during any arbitrarily chosen time period T does not exceed (b+rT)".

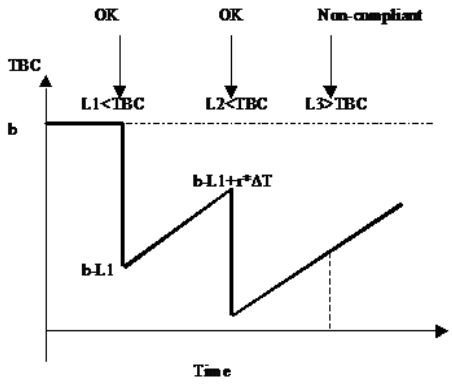


Fig. 2. Operation example of the reference conformance algorithm

The algorithm is described here (Figure 2, (3GPP23107, 2011)). Token bucket counter (TBC) is usually increased by "r" in each small time unit. However, TBC has upper bound "b" and the value of TBC shall never exceed "b". When a packet i with length Li arrives, the receiver checks the current TBC. If the TBC value is equal to or larger than Li, the packet arrival is judged compliant, i.e., the traffic is conformant. At this moment tokens corresponding to the packet length is consumed, and TBC value decreases by Li. When a packet j with length Lj arrives, if TBC is less than Lj, the packet arrival is non-compliant, i.e., the traffic is not conformant. In this case, the value of TBC is not updated.

The Maximum bitrate is the upper limit a user or application can accept or provide. All UMTS bearer service attributes may be fulfilled for traffic up to the Maximum bitrate depending on the network conditions. The downlink of the radio interface can use maximum bitrate to make code reservations. Its purpose is:

1. to limit the delivered bitrate to applications or external networks with such limitations;
2. to allow maximum wanted user bitrate to be defined for applications able to operate with different rates (e.g. applications with adapting codecs).

Guaranteed bitrate (kbps)

Guaranteed bitrate (kbps) is defined as: a guaranteed number of bits delivered by UMTS at a SAP within a period of time (provided that there is data to deliver), divided by the duration of the period. The traffic is conformant with the guaranteed bitrate as long as it follows a token bucket algorithm where token rate equals Guaranteed bitrate and bucket size equals Maximum SDU size.

UMTS bearer service attributes, e.g. delay and reliability attributes, are guaranteed for traffic up to the Guaranteed bitrate. For the traffic exceeding the Guaranteed bitrate the UMTS bearer service attributes are not guaranteed. Guaranteed bitrate may be used to facilitate admission control based on available resources, and for resource allocation within UMTS.

Delivery order (y/n)

Delivery order indicates whether the UMTS bearer shall provide in-sequence SDU delivery or not. This attribute is derived from the user protocol (PDP type) and specifies if

out-of-sequence SDUs are acceptable or not. Whether out-of-sequence SDUs are dropped or re-ordered depends on the specified reliability.

Maximum SDU size (octets)

Maximum SDU size (octets) means the maximum allowed SDU size. The maximum SDU size is used for admission control and policing.

SDU format information (bits)

SDU format information (bits) is a list of possible exact sizes of SDUs. UTRAN needs SDU size information to operate in transparent RLC protocol mode, which is beneficial to spectral efficiency and delay when RLC re-transmission is not used. Thus, if the application can specify SDU sizes, the bearer is less expensive.

SDU error ratio

SDU error ratio indicates the fraction of SDUs lost or detected as erroneous. SDU error ratio is defined only for conforming traffic. It is used to configure the protocols, algorithms and error detection schemes, primarily within UTRAN.

Residual bit error ratio Residual bit error ratio indicates the undetected bit error ratio in the delivered SDUs. If no error detection is requested, Residual bit error ratio indicates the bit error ratio in the delivered SDUs. It is used to configure radio interface protocols, algorithms and error detection coding.

Delivery of erroneous SDUs (y/n/-)

Delivery of erroneous SDU (yes/no/-) indicates whether SDUs detected as erroneous shall be delivered or discarded. 'Yes' implies that error detection is employed and that erroneous SDUs are delivered together with an error indication, 'no' implies that error detection is employed and that erroneous SDUs are discarded, and '-' implies that SDUs are delivered without considering error detection. It is used to decide whether error detection is needed and whether frames with detected errors shall be forwarded or not.

Transfer delay (ms)

Transfer delay (ms) indicates maximum delay for 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service, where delay for an SDU is defined as the time from a request to transfer an SDU at one SAP to its delivery at the other SAP. Transfer delay can be used to specify the delay tolerated by the application. It allows UTRAN to set transport formats and ARQ parameters.

Traffic handling priority

Traffic handling priority specifies the relative importance for handling of all SDUs belonging to the UMTS bearer compared to the SDUs of other bearers. Within the interactive class, there is a definite need to differentiate between bearer qualities. This is handled by using the traffic handling priority attribute, to allow UMTS to schedule traffic accordingly. By definition, priority is an alternative to absolute guarantees, and thus these two attribute types cannot be used together for a single bearer.

Allocation/Retention priority

Allocation/Retention priority specifies the relative importance compared to other UMTS bearers for allocation and retention of the UMTS bearer. The Allocation/Retention

Priority attribute is a subscription attribute which is not negotiated from the mobile terminal. Priority is used for differentiating between bearers when performing allocation and retention of a bearer. Where there is no enough resource, the relevant network elements can use the Allocation/Retention Priority to prioritize bearers with a high Allocation/Retention Priority over bearers with a low Allocation/Retention Priority when performing admission control.

Source statistics descriptor ('speech'/'unknown')

Source statistics descriptor ('speech'/'unknow') specifies characteristics of the source of submitted SDUs. Conversational speech has a well-known statistical behaviour (or the discontinuous transmission (DTX) factor). By using source statistics descriptor, a network element can know whether the SDUs of a UMTS bearer generated by a speech source or not. UTRAN, the SGSN and the GGSN and also the UE may, based on experience, calculate a statistical multiplex gain for use in admission control on the relevant interfaces.

2.3 QoS management for UMTS bearer service

In the section, an overview of QoS functions is described which is used to establish, modify, and maintain a UMTS bearer service with a specific QoS. The allocation of these functions to the UMTS entities indicates the requirement for specific entity to enforce the QoS commitments negotiated for the UMTS bearer service. UMTS is split into user plane and control plane for easy expanding in the future. So QoS management functions are also split into user plane and control plane. All of the QoS management functions in both planes (control and user plane) will ensure the provision of the negotiated service between the access points of the UMTS bearer service. The end-to-end service is provided by translation/mapping with UMTS external services.

2.3.1 QoS management in control plane

The QoS management functions in control plane are shown in Figure 3. The QoS functions for UMTS bearer service include service manager, translation function, admission/capability control and subscription control in the control plane. These functions are used to establish and modify a UMTS bearer service through signaling/negotiating with UMTS external services, establishing/modifying UMTS internal services.

Subscription control checks the administrative rights when an UMTS bearer service user requires a service with the specified QoS. It is located at CN EDGE.

Admission capability control will maintains available resource information of a network entity, and resource allocated to UTMS bearer service. When receiving an UMTS bearer service request or modification request, admission / capability control function determines whether the required resources can be provided or not. If the network can provide resources, it will reserve these resources. This function also checks the capability of a network entity, i.e. whether the specific service is implemented and not blocked for administrative reasons. It is located at MT, UTRAN, CN EDGE and Gateway.

Translation function is used to convert between internal service primitives and external protocols. It is located at MT and Gateway. At MT and Gateway, translation function converts between UMTS bearer service attributes and external network QoS attributes.

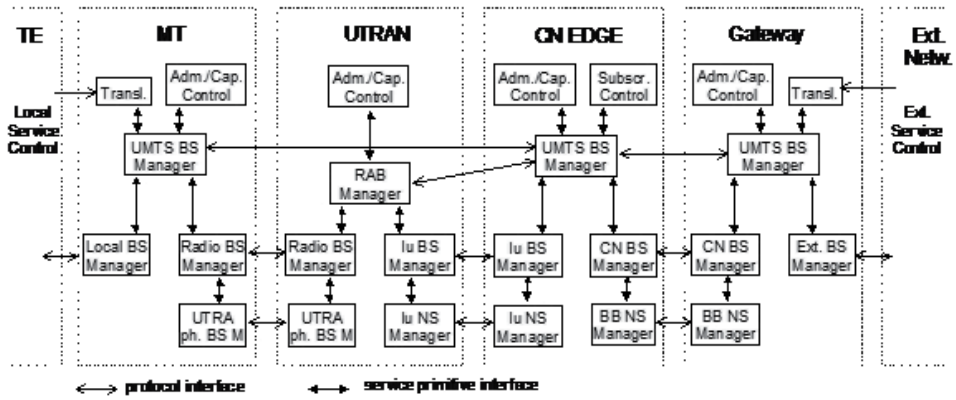


Fig. 3. QoS management function for UMTS bearer service in the control plane

Service manager co-ordinates the related functions in control plane to establish, modify and maintain the service. All user plane QoS management functions are supported by service manager with the relevant attributes. The service manager may perform an attribute translation to request lower layer services. Service manager at UMTS bearer service level is located at MT, CN EDGE and Gateway. The UMTS BS manager can signal among each other and via the translation function with external instances to establish / modify a UMTS bearer service. The UMTS BS manager will interrogate with its associated admission / capability control whether the network entity supports a specific requested service and whether the required resource is available. The UMTS BS manager at CN EDGE also has to verify with the subscription control the administrative rights for using the service. Based on the layered UMTS QoS architecture, UMTS bearer service manager will translate the UMTS bearer service attributes into attributes of the lower layer service manager. For example, the UMTS BS manager of the CN EDGE will translate the UMTS bearer service attributes into RAB service attributes, Iu bearer service attributes, and CN bearer service attributes. Each low layer will provide service to upper layer service manager.

2.3.2 QoS management in user plane

The Figure 4 shows the QoS management functions of UMTS bearer service in the user plane. They are mapping function, classification function, resource manager and traffic conditioner. They are used to maintain the data transfer characteristics according to the commitments established by the UMTS BS control functions.

Mapping function provides each data with the specific marking for receiving the requested QoS at the transfer. It is located at UTRAN, Gateway.

Classification function (Class.) in the MT and Gateway assigns user data units received from the external bearer service or the local bearer service to the appropriate UMTS bearer service according to the QoS requirement of each user data unit.

Traffic conditioner provides conformance between the negotiated QoS for a service nad the data unit traffic. Policing or traffic shaping is used for traffic conditioning. The policing function compares the data unit traffic with the related QoS attributes. Data units not matching the relevant attributes will be dropped or marked as not matching, for preferential

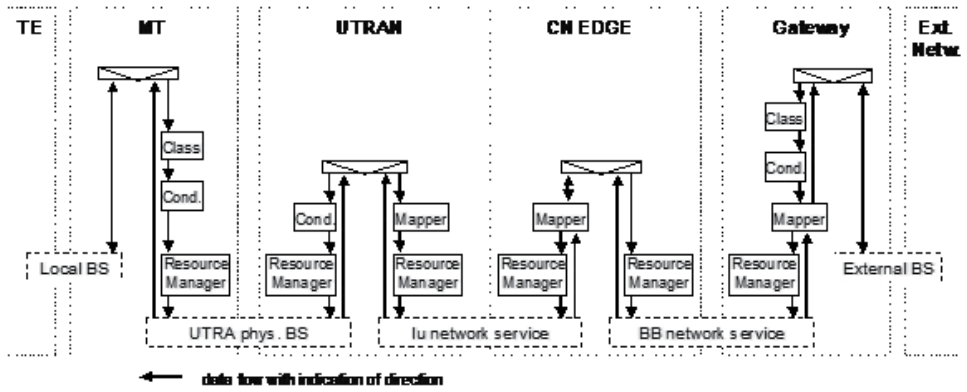


Fig. 4. QoS management function for UMTS bearer service in the user plane

dropping in case of congestion. The traffic shaper forms the data unit traffic according to the QoS of the service. The shaper algorithm is "Token Bucket Algorithm". At MT side, the traffic conditioner (Cond.) provides conformance of the uplink user data traffic with the QoS attributes of the relevant UMTS bearer service. In the Gateway a traffic conditioner may provide conformance of the downlink user data traffic with the QoS attributes of the relevant UMTS bearer service; i.e., on a per PDP context basis. A traffic conditioner in the UTRAN forms this downlink data unit traffic according to the relevant QoS attributes.

Resource Manager distributes the available resources between all services sharing the same resource. The resource manager distributes the resources according to the required QoS. Example means for resource management are scheduling, bandwidth management and power control for the radio bearer. It is located at MT, UTRAN, CN EDGE and Gateway.

2.4 QoS management in UMTS

2.4.1 QoS in CS domain

CS call control will control the QoS in the CS domain. MSC server and CS-MGW will provide QoS related functions. For UMTS release '99 CS-CC, the QoS related bearer definitions of GSM (as defined in bearer capability information element, octet 6 and its extensions) are sufficient.

In the CS domain the UE can only request a certain service with a well defined set of QoS parameters. CS domain uses traditional "circuit switching" technology, i.e. a constant set of resources exclusively dedicated to a connection. All CS domain services will require real-time bearers but differ in bandwidth and delay requirements. Based on the Bearer Capability information element the following services can be identified:

1. speech: from the Information Transfer Capability (ITC) parameter;
2. data, non-transparent: from the ITC and Connection element (CE) parameters;
3. data, transparent: from the ITC and CE parameters.

According to the standard, speech as well as the transparent data service is mapped to the conversational class while the non-transparent data service is mapped to the streaming class.

The MSC-Server is responsible for the service negotiation which includes subscription check and admission control. Furthermore, the QoS parameters corresponding to the service have

to be mapped specifically for the interfaces to the UTRAN, GMSC-Server and CS-MGW. To provide QoS the CS-MGW has to perform admission control for the bearer resource which is therefore a part of the call admission control. Additionally, the CS-MGW is responsible for the QoS mappings to the Iu-, CN- and external bearer services.

With the separation of transport and control in the CS domain the resource allocation becomes more flexible. The new transport techniques ATM and IP (which are available for the CS bearer independent domain) allow a more efficient network usage from a parallel transmission of voice and data possibly leading to the consolidation of the whole PLMN (including the PS domain and parts of RAN) on one transport network. The QoS issue in the CS domain with IP or ATM based transport is to guarantee the same QoS as a TDM based PLMN with increased bandwidth efficiency.

2.4.2 QoS in PS domain

Since the PS domain provides packet data services, which are characterized by individual transmission of packets. QoS of different packet service is defined by a set of explicitly defined QoS parameters. So some effort is necessary to assure that packets of one flow are transmitted with guaranteed QoS.

The 3GPP specificationscitep (3GPP23107, 2011) define the QoS management functions in the UMTS bearer service for both control plane and user plane. Establishment of QoS within a UMTS network is achieved through the Packet Data Protocol (PDP) context activation procedure. The user equipment (UE) sends an Active PDP Context Request message to the SGSN, which contains the desired QoS profile, among other parameters. With these QoS attributes the treatment of the packets is sufficiently defined and all packets (or flows) belonging to the same PDP context are handled in the same way by the GPRS bearer service. After the UE sends a PDP context request with explicitly defined QoS parameters, the SGSN will negotiate the QoS parameters which includes subscription check and admission control (capability and resource check). Then the SGSN interacts with the UTRAN and the GGSN to establish the PDP context. The GGSN also performs admission control, i.e. the resource check for the GPRS as well as for the external bearer service. Additionally, the GGSN has to map QoS parameters from the GPRS to the external bearer service.

2.4.3 QoS in IP multimedia subsystem

IP multimedia subsystem (IMS) is introduced in 3GPP Release 5. It is an IP based system overlay on the PS domain. It support Session Initiation Protocol (SIP) based multimedia service. IMS can support end-to-end IP QoS service by using IP based bearer service. The IP based bearer service is supported by MS local bearer service, UMTS bearer service and external service.

Since 3GPP Release 5, the UMTS will support QoS in the IP layer between UE and multimedia application server/UE. The UE and GGSN have important roles in the IP layer QoS framework, they map QoS parameters between IP layer bearer service and UMTS bearer service. The detail will be discussed in the section 3.

For supporting IP layer QoS, 3GPP introduces the policy based QoS management in the IMS. The policy framework is recommended for policy management in IETF. The detail discusses is given in the section 4.

3. End-to-end IP QoS over UMTS

With the evolution of the 3GPP standards, operators want to provide end-to-end QoS enabled services in UMTS. The end-to-end behavior provided by a series of network elements is an assured level of bandwidth that produces a delay-bounded service with no queueing loss for all conforming packet data (RFC2212, 1997). Assuming the network is functioning correctly, these applications may assume that (**?**):

- A very high percentage of transmitted packets will be successfully delivered by the network to the receiving end-nodes. (The percentage of packets not successfully delivered must closely approximate the basic packet error rate of the transmission medium).
- The transit delay experienced by a very high percentage of the delivered packets will not greatly exceed the minimum transmit delay experienced by any successfully delivered packet. (This minimum transit delay includes speed-of-light delay plus the fixed processing time in routers and other communications devices along the path.)

The end-to-end QoS architecture is provided in Figure 1 in section 2. IP level mechanisms are necessary in providing end-to-end QoS services by interacting TE/MT local bearer service, GPRS bearer service and external bearer service. In this section, how to implement end-to-end IP QoS is described.

3.1 QoS mechanisms in IP

Quality of service refers to the nature of the packet delivery service provided, as described by parameters such as achieved bandwidth, packet delay, and packet loss rates (RFC2216, 1999). The Internet, as originally conceived, offers only a very simple quality of service (QoS), point-to-point best-effort data delivery. It means the network just offered available bandwidth and delay characteristics dependent on instantaneous network load. Before real-time applications such as remote video, multimedia conferencing, visualization, and virtual reality can be broadly used, the Internet infrastructure must be modified to support real-time QoS, which provides some control over end-to-end packet delays. From the view of applications, QoS is realized by adequate provisioning of the network infrastructure. In contrast, a network with dynamically controllable quality of service allows individual application sessions to request network packet delivery characteristics according to their perceived needs, and may provide different qualities of service to different applications. There are two basic types of QoS available (qodwhitepaper, 1999):

- Resource reservation (integrated services): network resources are apportioned according to an application's QoS requirement, subject to bandwidth management policy.
- Prioritization (differentiated services): network traffic is classified and apportioned network resources according to bandwidth management policy.

The both types of QoS can be applied to individual application 'flow' or to flow aggregates, so there are two other methods to characterize types of QoS:

- Per flow: A 'flow' is defined as an individual, uni-directional data stream between two clients (caller and callee), uniquely identified by a 5-tuple (transport protocol, source address, source port number, destination address, and destination port number).
- Per aggregate: An aggregate is simply two or more flows. Usually the flows have something in common (e.g. any one or more of 5-tuple parameters, a label or a priority number, or perhaps some authentication information).

Generally, we can see that there are two methods to support QoS in IP network. One is IntServ (Integrated Service), the other is DiffServ (Differentiated Service). IneServ is Per Flow based QoS control mechanism. Diffserv is Per Aggregate based QoS control mechanism. To accommodate the need for these two types of QoS, there are following QoS protocols and algorithms:

- ReSerVation Protocol (RSVP):
- Differentiated Service (DiffServ)
- Multi Protocol Labeling Switching (MPLS)

3.1.1 IntServ

The Internet integrated services (IntServ) framework provides the ability for applications to choose among multiple, controlled levels of delivery service for their data packets. It can provide hard QoS guarantee to individual traffic flows. To support this capability, two things are required (**?**):

- Individual network elements (subnets and IP routers) along the path followed by an application's data packets must support mechanisms to control the quality of service delivered to those packets.
- A way to communicate the application's requirements to network elements along the path and to convey QoS management information between network elements and the application must be provided.

In the integrated services framework the first function is provided by QoS control services such as Controlled-Load (RFC2211, 1997) and Guaranteed (RFC2212, 1997). The second function may be provided in a number of ways, but is frequently implemented by a resource reservation setup protocol such as RSVP (RFC2205, 1997).

The controlled load service is intended to support a broad class of applications which have been developed for use in today's Internet, but are highly sensitive to overloaded conditions. Important members of this class are the ”adaptive real-time applications” currently offered by a number of vendors and researchers. These applications have been shown to work well on unloaded nets, but to degrade quickly under overloaded conditions. It is equivalent to ”best effort service under unloaded conditions”. The controlled-load service is intentionally minimal, in that there are no optional functions or capabilities in the specification. The service offers only a single function. It is better than best effort, but cannot provide strictly bounded service as guaranteed service.

The controlled-load service can be implemented by using evolving scheduling and admission control algorithms. The implementations are highly efficient in the use of network resources.

Guaranteed service guarantees that datagrams will arrive within the guaranteed delivery time and will not be discarded due to queue overflows, provided the flow's traffic stays within its specified traffic parameters. It is similar to emulate a dedicated virtual circuit. This service is intended for applications which need a firm guarantee that a datagram will arrive no later than a certain time after it was transmitted by its source. For example, some audio and video "play-back" applications are intolerant of any datagram arriving after their play-back time. Applications that have hard real-time requirements will also require guaranteed service.

Guaranteed service does not attempt to minimize the jitter (the difference between the minimal and maximal datagram delays); it merely controls the maximal queueing delay. Because the guaranteed delay bound is a firm one, the delay has to be set large enough to cover extremely rare cases of long queueing delays. Several studies have shown that the actual delay for the vast majority of datagrams can be far lower than the guaranteed delay. Therefore, authors of playback applications should note that datagrams will often arrive far earlier than the delivery deadline and will have to be buffered at the receiving system until it is time for the application to process them.

Guaraneteed service represents one extreme end of delay control for networks. Most other services providing delay control provide much weaker assurances about the resulting delays. In order to provide this high level of assurance, guaranteed service is typically only useful if provided by every network element along the path (i.e. by both routers and the links that interconnect the routers). Moreover, as described in the Exported Information section, effective provision and use of the service requires that the set-up protocol or other mechanism used to request service provides service characterizations to intermediate routers and to the endpoints.

Integrated Services routers uses admission control and resource allocation method to offer QoS guarantee. A token-bucket model is used to characterize the input/output queueing algorithm. It can smooth the flow of outgoing traffic. The IntServ parameters include (qodwhitepaper, 1999):

Token rate (r): The continually sustainable bandwidth (bytes/second) requirement for a flow. It represents the average data rate into the bucket, and the target shaped data rate out of the bucket.

Token-bucket rate (b) : the extent to which the data rate can exceed the sustainable average for short periods of time, or the amount of data sent cannot exceed rT+b (where T is any time period).

Peak rate (p): It is the maximum send rate (bytes/second) if known and controlled. At any time period (T), the amount sent data cannot exceed M+pT.

Minimum policed size (m): The size (byte) of the smallest packet (data payload only) can be generated by the sending application. The size m is not an absolute number. If the percentage of small packets is small, the number m should increased to reduce the overhead estimate. All packets smaller than m are treated as size m.

Maximum packet size (M): The biggest size of a packet (bytes). The M is absolute number. Any packets (size > M) are considered out of spec and may not receive QoS controlled service.

3.1.2 RSVP

For offering IntServ, a way to communicate the application's requirements to network elements along the path and to convey QoS management information between network elements and the application must be provided. A resource reservation setup protocol called RSVP (rfc2205, 1997) is implemented for this purpose. It is a signaling protocol that can provide reservation setup and control to enable the integrated services by using a variety of QoS control, a variety of setup mechanisms.

A simplified RSVP working flow is shown in Figure 5

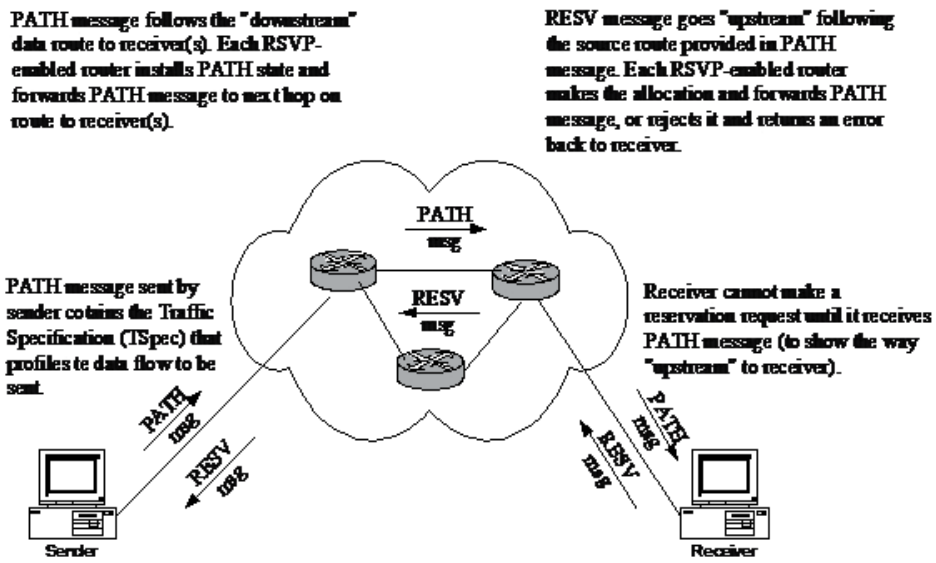


Fig. 5. RSVP setup flow

1. When a sender wants to set up a traffic link, it will generate the traffic specification (TSpec), which describes data traffic, such as upper/lower bounds of bandwidth, delay, and jitter. Then RSVP sends out a PATH message containing TSpec to the receiver(s) (unicast or multicast). Along the route, each RSVP-enabled router trigger a "path-state" that includes the previous source address of the PATH message.
2. After receiver receives the PATH message, the receiver sends a RESV message "upstream" to make a resource reservation. The RESV message includes a request specification (RSpec) which indicates what type of IntServ required âĂŞ either Controlled Load or Guaranteed, a filter specification (filter spec) (indicating e.g. the transport protocol and port number). The RSpec and filter spec represent a flow-descriptor that RSVP routers use to identify each reservation.
3. Along the RSVP upstream, RSVP routers use the admission control to authenticate the resource reservation request and allocate the necessary resources when the routers receive the RESV message. If the request cannot be met (due to no enough bandwidth or authorization failure), the RSVP router returns an error back to the receiver. If the request is accepted, the router forwards the RSVP message to the next router.
4. When the last router receives the RESV message and accepts the request, it sends out a confirmation message back to the receiver.
5. There is an explicit tear-down process for a reservation when sender or receiver terminate a RSVP session.

3.1.3 DiffServ

The Integrated Services/RSVP model relies upon traditional datagram forwarding in the default case, but allows sources and receivers to exchange signaling messages which establish additional packet classification and forwarding state on each node along the path between

them (rfc1633, 1994). In the absence of state aggregation, the amount of state on each node scales in proportion to the number of concurrent reservations, which can be potentially large on high-speed links. This model also requires application support for the RSVP signaling protocol. Differentiated service is a simple method by classifying services of different applications (rfc2475, 1998). Currently there are two standard per hop behaviour (PHBs) define two traffic classes:

- Expedited Forwarding (EF): Has a single codepoint (DiffServ value). Ef minimize delay and jitter and provides the highest level of aggregate quality of service. Any traffic that exceeds the traffic profile is discarded (**?**). EF class offers a low jitter, low delay service. UserâĂŹs traffic cannot exceed the agreed peak rate. Otherwise, the packets will be discarded.
- Assured Forwarding (AF): Has four classes and three drop-precedence within each class (a total of twelve codepoints). Excess AF traffic is not delivered with as high probability as the traffic "within profile", which means it may be demoted but not necessarily dropped (**?**). The AF class is suitable for delay-tolerant applications. The guarantee just implies that the better QoS class will give a better performance than the low-level QoS class. Network operator can define their own per-hop behavior.

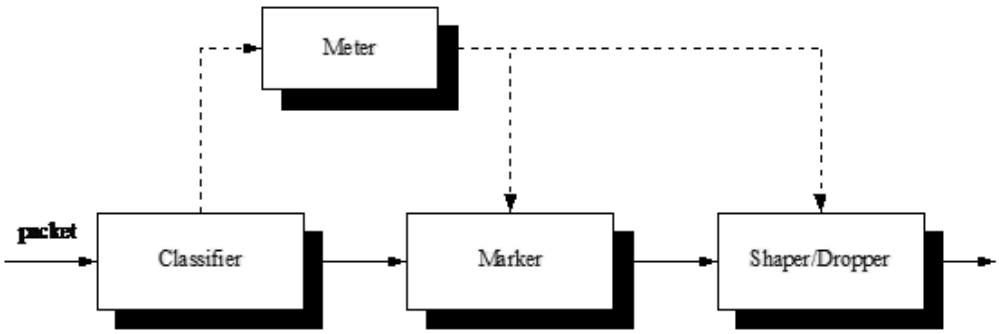


Fig. 6. DffServ architecture

DiffServ offers a simple QoS management method without signaling mechanism. The DiffServ architecture is shown in Figure 6. It includes classifier and traffic conditioner. A traffic conditioner contains the following elements: meter, maker, shaper/dropper. The differentiated services architecture is based on a simple model where traffic entering a network is classified and possibly conditioned at the boundaries of the network, and assigned to different behavior aggregates. Each behavior aggregate is identified by a single DiffServ codepoint (DSCP). Within the core of the network, packets are forwarded according to the per-hop behavior associated with the DiffServ codepoint.

When a traffic flow enters a DiffServ network, the flow is selected by a classifier, which steers the packets to a logical instance of a traffic conditioner. A meter is used to measure the traffic flow agains a traffic profile. A meter is used (where appropriate) to measure the traffic stream against a traffic profile. The state of the meter with respect to a particular packet (e.g., whether it is in- or out-of-profile) may be used to affect a marking, dropping, or shaping action. When packets exit the traffic conditioner of a DS boundary node the DiffServ codepoint of each packet must be set to an appropriate value.

3.1.4 MPLS

MPLS is a key development in IETF that will add a number of essential capabilities to today's best effort IP networks, including

- Traffic Engineer, enhancing overall network utilization by creating a uniform or differentiated distribution of traffic throughout the network.
- Providing traffic with different Classes of Service (CoS)
- Providing traffic with different Quality of Service (QoS)
- Supporting network scalability, providing IP based Virtual Private Networks (VPN)

MPLS borrows the idea from ATM switching. It remains independent of the Layer-2 and Layer-3 protocols. Besides IP, other network protocols (such as IPX, ATM, PPP or Frame-Relay) also can work with MPLS. MPLS resides on routers. When a packet flow enters a edge router of the MPLS domain, all packets are marked to clarify priority with a fixed-length label (20 bits label). The label identifies the packets routing information in this MPLS network, also define the quality of service for the packets.

A MPLS domain includes label edge routers (LERs) and label switching routers (LSRs). The route taken by an MPLS-labeled packet is called the label switched path (LSP). LST is a high-speed router in the core of a MPLS network, which participates in the establishment of LSPs. LER is a router that operates at the edge of a MPLS network. It is used to assign and remove labels when packets enter or exit the MPLS network.

MPSL is similar to DiffServ because it also marks traffic at ingress of a MPLS network, and un-marks at egress gate. However, MPLS marking is used to decide the next hop router while DiffServ marking is used to determine priority in route itself.

3.2 QoS management functions for end-to-end IP QoS

This section describes how to provide Quality of Service in UMTS for the end-to-end services through the TE/MT local bear service, GPRS bearer service and external bearer service shown in the Fig. 1. To provide end-to-end IP QoS, it is necessary to manage the QoS within each domian. An IP BS Manager is used to control the external IP bearer service. Due to the different techiniques used within the IP network,this communicates to the UMTS BS manager through the Translation function.

At PDP context setup the user shall have access to one of the following alternatives, basic GPRS IP connectivity service or enhanced GPRS based services. To enable coordination between events in the application layer and resource management in the IP bearer layer, a logical element, the Policy and Charging Rules Function (PCRF), is used as a logical policy decision element which will be detailed in section 4. It is also possible to implement a policy decision element internal to the IP BS Manager in the GGSN. While interworking with the external network, the RSVP, DiffServ, MPLS will be used.

QoS management functions is shown in Fig. 7 which describes how to control the external IP bearer services and how they relate to the UMTS bearer service QoS management entity.

IP BS Manager uses standard IP mechanisms to manage the IP bearer services. These mechanisms may be different from mechanisms used within the UMTS, and may have different parameters controlling the service. When implemented, the IP BS Manager

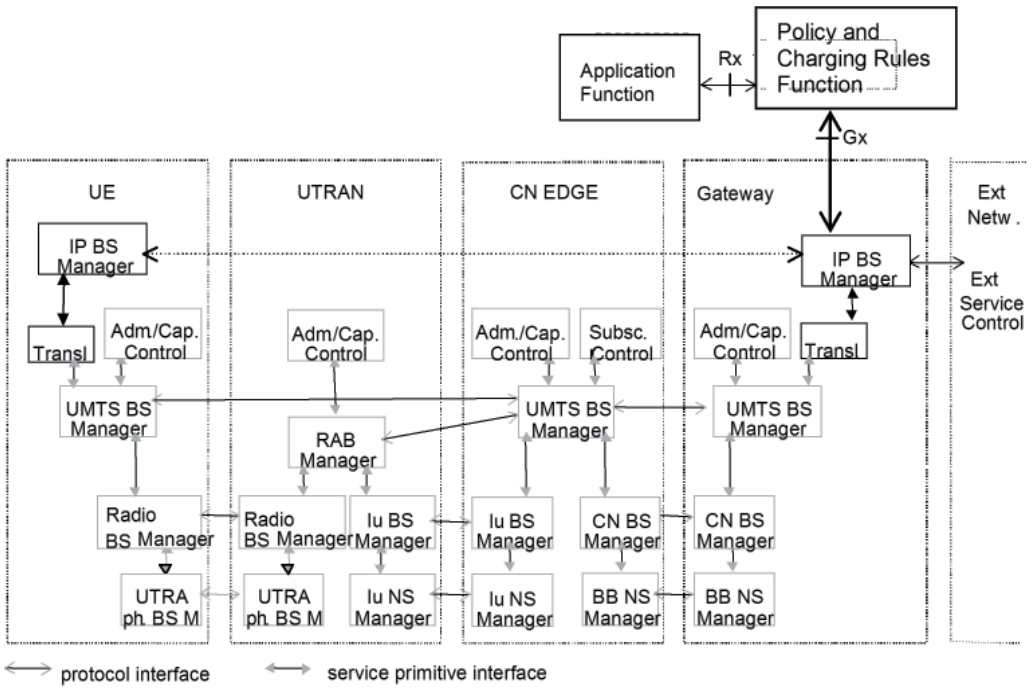


Fig. 7. QoS management functions for end-to-end QoS in UMTS

may include the support of DiffServ Edge Function and the RSVP function. The Translation/mapping function provides the inter-working between the mechanisms and parameters used within the UMTS bearer service and those used within the IP bearer service, and interacts with the IP BS Manager. In the GGSN, the IP QoS parameters are mapped into UMTS QoS parameters, where needed. In the UE, the QoS requirements determined from the application layer (e.g., SDP) are mapped to either the PDP context parameters or IP layer parameters (e.g., RSVP). If an IP BS Manager exists both in the UE and the Gateway node, it is possible that these IP BS Managers communicate directly with each other by using relevant signalling protocols. The required options in the table define the minimum functionality that shall be supported by the equipment in order to allow multiple network operators to provide interworking between their networks for end-to-end QoS. Use of the optional functions listed below, other mechanisms which are not listed (e.g. over-provisioning), or combinations of these mechanisms are not precluded from use between operators. The IP BS Managers in the UE and GGSN provide the set of capabilities for the IP bearer level as shown in table 2. Provision of the IP BS Manager is optional in the UE, and required in the GGSN.

Capability	UE	GGSN
DiffServ Edge Function	Optional	Required
RSVP/IntServ	Optional	Optional
IP Policy Enforcement Point	Optional	Required

Table 2. IP BS Manager capability in the UE and GGSN

4. Policy based QoS management - IP QoS for IMS

This section will provide an overview of policy based QoS management in UMTS IMS. Although the UMTS packet switched (PS) domain can support IP QoS enabled multimedia applications, there are many ways of establishing QoS guaranteed IP multimedia session through a signaling protocol before it can map and reserve the equivalent amount of QoS resources along the data path in the PS domain. In order to support interoperation among UMTS network providers, the IP multimedia Subsystem (IMS) is standardized by the 3GPP to serve Session Initiation Protocol (SIP) signaled IP multimedia services over the UMTS PS domain.

The central problem of providing consistent end-to-end IP QoS services is the difficulty of configuring the network devices like routers and switches to handle packet flows in a manner that satisfies the requested QoS requirements. Policy-based QoS management is used to control QoS resources in the UMTS IMS.

4.1 Introduction to policy-based QoS network

4.1.1 The need for policy based network

There is a consistent effort to implement new IP multimedia services in UMTS. While the IP based network is well suited for packet data transfer, providing consistent end-to-end IP QoS services is the difficulty of configuring the network devices like routers and switches to handle packet flows in a manner that satisfies the requested QoS requirements. This problem is especially acute when the end-to-end data path of an IP QoS session crosses multiple administrative domains managed by different operators. Although the operators agree on the QoS requirements of a particular set of IP services, they may not configure their network devices in the same way to implement the services due to differences in the network topologies, QoS mechanisms available in the network devices and non-technical management requirements. Thus, there is a need to create a solution that permits network operators, including UMTS network operators, to easily configure their networks to implement consistent IP QoS services without dealing with the complexity of their networks.

Policy-based Networking (PBN) is a novel approach to configure myriad network devices in an administrative domain to implement a set of IP QoS services. Policy-based network will allow the network operator to define, in a succinct and organized fashion, operator policies that automatically effect change on specific equipment in the network environment. The end result is that the end-to-end network performance will meet the general expectations of UMTS service provider environment.

4.1.2 What is policy?

A policy is a set of business rules that guide and determine how to manage network resources. The basic concept is that policy rule(s) describe how network to act when specific condition(s) happen. "Policy" can be defined from two perspectives: (POLICYTERM, 2001). - A definite goal, course or method of action to guide and determine present and future decisions. "Policies" are implemented or executed within a particular context (such as policies defined within a business unit). - Policies as a set of rules to administer, manage, and control access to network resources. [RFC3060] Note that these two views are not contradictory since individual rules may be defined in support of business goals.

Policy can be represented at different levels, ranging from business goals to device-specific configuration parameters. Enforcement of policy ensures that business rules are always followed. Policy rule is a basic building block of a policy-based system. It is the binding of a set of actions to a set of conditions - where the conditions are evaluated to determine whether the actions are performed. [RFC3060] A condition is a set of expressions or objects used to determine whether a given policy rule's action should be performed. A condition answers the question, "when and where do we enforce a policy?" An action defines what to be done to enforce a policy rule, when the conditions of the rule are met. Policy actions may result in the execution of one or more operations to affect and/or configure network traffic and network resources. An action answers the question, "what must be done to enforce a policy?"

A policy also defines how the network's resources are to be allocated among its clients. Clients can be individual users, departments, host computers, or applications. Resources can be allocated based on time of day, client authorization priorities, availability of resources, and other factors. How resources are allocated can be static or dynamic (based on variations in traffic). Policies are created by network managers and stored in a repository. During network operation, the policies are retrieved and used by network management software to make decisions.

4.1.3 Policy framework & architecture

The network operators negotiate Service Level Agreements (SLAs) that describe the sets of IP QoS services that they have mutually contracted to provide. Individual operators will then transform the QoS requirements specified in the SLAs into sets of policy rules that will be applied to their network domains to implement the contracted IP QoS services. The IETF has defined a policy framework (RFC2753, 2001) as shown in Figure 8 to transform the sets of policy rules to network device configurations in an administrative domain. The sets of policy rules are stored in the Policy Repository through the Policy Management Tool. The Policy Decision Point (PDP) retrieves the appropriate policy rules from the Policy Repository in response to policy events that are triggered by the contracted IP QoS services, e.g., the reception of an RSVP message by the Policy Enforcement Point (PEP). It translates the acquired policy rules into a set of QoS mechanism configuration actions that is communicated to the PEP as policy decisions. The PEP then executes the actions spelt out in the supplied decisions to handle the triggering policy events in accordance with the requested IP QoS services. Alternatively, the retrieved policy rules may be returned to the PEP, which is capable of translating them into configuration actions. These policy rules can be cached in the PEP so that similar future triggering policy events can be serviced locally without further interactions with the PDP.

Outsourcing and Provision Model in PBN

There are two main models for policy management: outsourcing and provisioning. The outsourcing model assumes there is a signaled event in the Policy Enforcement Point (PEP) that must be resolved based on policy criteria. The PEP outsources the decision-making to an external policy decision point (PDP). This outsourcing model is sometimes referred to as "Pull" mode, or "reactive" mode, since the PEP pulls policy decisions from the PDP, while the PDP responds according to the PEP events.

The provisioning model is almost the mirror image of the outsourcing model. In this system, the PDP predicts future configuration needs, and proactively provisions resources accordingly. In other words, rather than responding to PEP events, the PDP prepares

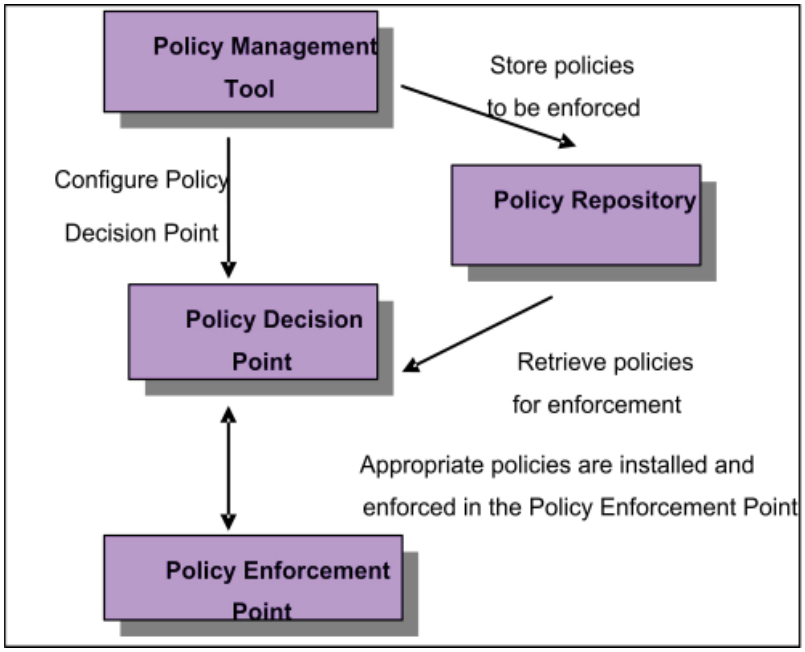


Fig. 8. A PBN architecture that is derived from the policy framework specified by the IETF

and "pushes" configuration information to the PEP. This takes place as a result of external events (unrelated to the PEP) such as change of applicable policy, time of day, expiration of account quota, or information from third party (non-PEP) signaling.

Both models employ policy servers as the PDP to control the network devices that enforce the policy (i.e. PEPs). PBN also offers a policy repository for storing policy information accessed by the PDPs in the system. To communicate policy information between PDPs and PEPs, the COPS policy protocol is engaged. Additionally, the LDAP protocol functions to access the policy repository.

Policy Decision Point (PDP)

The PDP is the PBN component that directly controls the network devices or policy enforcement points (see next section). Functionally, the PDP handles policy information that has been entered into the PBN management system. The policy data used by the PDP can either be obtained in real-time upon entry into the management console, or from the policy repository on an as-needed basis. The function of the PDP involves retrieving policy, interpreting policy, detect policy conflicts, receiving policy decision requests from PEPs, determining which policy is relevant, applying the policy and returning the results. It also sends policy elements the PEP.

Policy Enforcement Point (PEP)

Network devices that receive and enforce the decisions from the PDP are referred to as PEPs. In both outsourcing and provisioning policy management models, PEPs receive policy decisions and enforce them at the packet level as data passes through the devices.

4.2 Policy framework in UMTS IMS

To support IP based multimedia services, the IP Multimedia Subsystem (IMS) is introduced in the 3GPP Release 5 specifications. It provisions IP based multimedia services as an extension of the UMTS PS domain (Figure 9). The added IMS functionalities are control functionalities; the user data traffic is still carried by the PS domain. The main advantage of the IMS is that it offers operators a scalable service platform on which new services can be developed rapidly in a flexible way, without requiring any change to the PS domain.

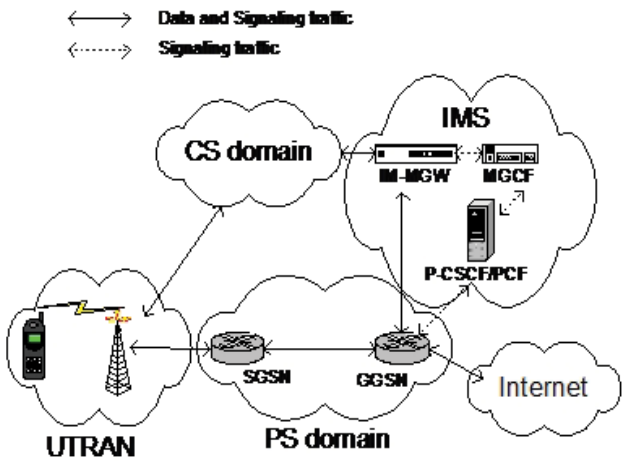


Fig. 9. A simple UMTS network with IMS

Having put in place the functionalities to handle IP multimedia calls, the next big challenge is to ensure that sufficient QoS resources are provided to authorized users in the UMTS network. A policy-based QoS solution is adopted by the 3GPP for this purpose.

As mentioned in section 4.1.3, the reference model of a policy-based network consists of two main elements, the PDP and the PEP (RFC2753, 2001). PEPs often reside in policy aware network nodes that carry out actions stipulated by policy rules. The actions taken are based on the decisions of a PDP, which retrieves the policy rules from a repository. The PDP is the final authority, which the PEP needs to refer to for actions to be taken.

In the IMS, the Policy and Charging Rules Function (PCRF) (3G23203, 2008) plays the role of the PDP and online charging and offline charging functions, the Policy and Charging Enforcement Functions plays the role of the PEP. Policy charging and rules function (PCRF) is the node designated in real-time to determine policy rules in a multimedia network. As a policy tool, the PCRF plays a central role in WCDMA networks. Unlike earlier policy engines that were added on to an existing network to enforce policy, the PCRF is a software component that operates at the network core and efficiently accesses subscriber databases and other specialized functions, such as a charging systems, in a scalable, reliable, and centralized manner. The PCRF as the part of the network architecture that aggregates information to and from the network, operational support systems, and other sources (such as portals) in real time, supporting the creation of rules and then automatically making intelligent policy decisions for each subscriber active on the network. Such a network might offer multiple services, quality of service (QoS) levels, and charging rules. In this chapter, we will focus on policy based management functions.

The PCRF communicates with the PCEF via the Gx interface (3G29212, 2008). It allows two modes of operation. In the "push" mode, the PCRF initiates communication with the PCEF and sends the PCEF its decision. In the "pull" mode, the PCEF initiates communication with the PCRF to request a decision for a particular IP flow. The Gx interface and the protocol used for communication on the interface are described in the following.

Figure 10 depicts the relationship between these entities. In the following subsections, each of these network elements will be described.

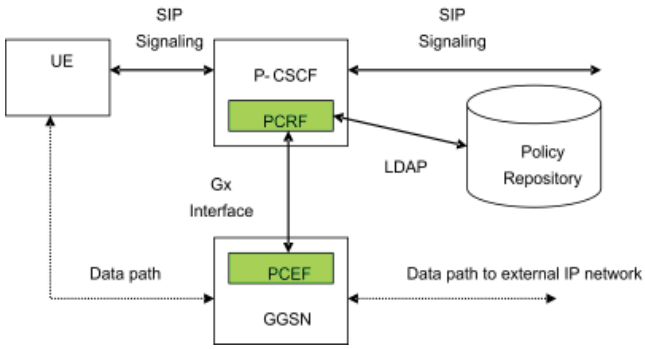


Fig. 10. Policy architecture in UMTS IMS

PCRF in Proxy-CSCF

During the establishment of a SIP session, a P-CSCF is the first contact point in the IMS domain for a UE [3G23228]. Hence it is the natural place to authorize the usage of network resources such as the bandwidth requested by the UE. The QoS requirements of the UE are carried in the Session Description Protocol (SDP) description within a Session Initiation Protocol (SIP) message. Besides the QoS requirements in the SDP description, the PCRF also examines the source and destination IP addresses and port numbers in its decision-making. The PCRF refers to the policy rules, which are generally stored in a policy repository, governing the local domain. It then generates an authorization token that uniquely identifies the SIP session across multiple Packet Data Protocol (PDP) contexts terminated by a GGSN. This token is sent to the UE via SIP messages so that the UE can use it to identify the associated session flows to the PCEF in the GGSN in subsequent transmission of IP packets. This mechanism is consistent with the IETF specification on supporting media authorization in the SIP protocol [RFC3313]. The flow of events in session set-up is described in Section 4.6.

PCEF in Gateway GPRS Support Node (GGSN)

In the PS domain, a GGSN maintains connectivity to other packet switched external networks such as the Internet. From the service point of view, the GGSN controls which IP flows are permitted into the external IP network by policing the IP packets based on their source and destination IP addresses and port numbers [3G23228]. As such, it is logical to embed the PCEF in the GGSN. The role of the PCEF is to ensure that only authorized IP flows are allowed to use network resources that have been reserved and allocated to them.

The policy enforcement function in the GGSN is called a "gate". A gate comprises a packet classifier, a traffic meter, and the relevant packet handling mechanisms for packets that have been matched by the packet classifier. When an IP flow is authorized by the PCRF to

use the specified network resources, the PCEF opens the "gate" for the flow and effectively commits the network resources to the flow by allowing it to pass through the packet handling mechanisms (i.e., policing or marking). On the other hand, if an IP flow is not permitted by the PCRF to use the requested resources, the PCEF closes the âĂIJgateâĂİ and drops the IP packets of the flow. This process is called policy-based admission control. It ensures that an IP flow is only allowed to use resources that have been approved by the policy rules. The above process takes place at the IP bearer service (BS) level. The translation/mapping function within the GGSN will map this resource information into the format used by the admission control function at the UMTS BS level.

The PCEF may store decisions in a local policy decision point, thus allowing the GGSN to make the admission control decisions without additional interactions with the PCRF. This will reduce the traffic over the Gx interface and lessen the processing load on the PCRF.

4.3 Policy-based QoS delivery: an example of policy based call control

There are several reasons why a policy-based QoS framework is adopted for the UMTS. Policy-based QoS control allows network operators to configure their network devices easily. It provides a high level view of the network devices and allows the automated translation of business level policies to suitable information for configuring network devices.

UMTS requires a strict authorization of users so that the network resources are not abused. Once authorized and approved, the UMTS must guarantee that the resources are made available to the legitimate users. If these requirements are not met, these users may be denied the use of the resources, leading to dissatisfaction with the quality of service provided. To ensure that this is not the case, all IP multimedia calls must go through the following steps:

1. Authorization of resources;
2. Reservation of resources. This is to make sure that the resources are available when the "phone" rings;
3. Once the called party picks up the "phone", the network resources reserved previously are committed. The charging process is then triggered.

In all these steps, policy rules are used in approving the requests, and the PCRF is the sole approving authority. By changing the policy rules in the PCRF, a network operator can alter the IP multimedia services it offers to its subscribers without having to know the details of its network configuration and the types and mechanisms of the network devices.

To meet the above requirements, two procedures are needed for the establishment of an IP multimedia session in addition to the normal GPRS bearer establishment procedures. These procedures are Authorize QoS Resources and Approval of QoS Commit (3G29212, 2008). Similarly, the procedures, Removal of QoS Commit and Revoke Authorization of QoS Resources, are carried out to reverse the authorization and commitment of QoS resources when an IP multimedia session is terminated. The following provides an overview of the session set-up procedures, in particular, the emphasis is on the additional procedures introduced by the service-based local policy.

4.4 Session establishment procedures

The establishment of an IP multimedia service session with policy control differs from that without policy control in that additional steps are taken to check the policy rules for a decision

on whether to grant or deny the required network resources to the session. As the signaling messages used to set up the session take a different path from that used for the data flow, an authorization token and a flow identifier are used to associate the session with its IP data flow (UMTSGO01, 2001). The GGSN, which is located on the data path, relies on this binding information to enforce the policy rules on the IP data flows.

Figure 11 depicts the sequence of events that take place during the establishment of an IP multimedia service session. Note that a number of signaling messages have been omitted for clarity. The events are described in the following paragraphs:

Steps 1-5: The UE sends a session set-up request (i.e., SIP INVITE) to the P-CSCF indicating, among other things, the media streams to be used in the session. This message is routed to the called party via a number of other CSCFs (viz., the caller and callee S-CSCFs) along the signaling path. The S-CSCFs perform the appropriate session control services for the UEs. In particular, they maintain a session state that is needed by the network operator to support the requested service.

Steps 6-14: The called party responds with a provisional SIP 183 response message. This message is routed to the calling party via the same CSCFs along the (reversed) signaling path. When the callee P-CSCF receives this message, it examines the SDP description within the message to determine the QoS parameters requested for the session. The P-CSCF sends the necessary information in this SIP message (e.g., the bandwidth, IP addresses and ports, etc.) to the PCRF for authorization of the session request. If the policy permits, the PCRF responds with an authorization token that can be used to identify the authorized session and resources. The P-CSCF includes the token in the response (SIP 183 message) and forwards it to the callerâĂŹs UE. A similar process is carried out at the caller P-CSCF when it receives the SIP 183 message. This process of authorization by the PCRF and the generation of a token is called "Authorize QoS Request".

Steps 15-22: In between steps 14 and 15, other message exchanges take place between the caller and the callee. However, these are not important in this particular example and are omitted for clarity. The caller's UE starts the resource reservation by sending a PDP Context Activation Request to the GGSN. The authorization token and the flow identifier(s) from the PCRF are included to identify the IP data flow(s) of the session. When the GGSN receives the PDP Context Activation Request, it sends a policy decision request to the PCRF to determine whether the resource reservation request should be accepted. The PCRF uses the token in the message to correlate the request for resources with the media authorization previously granted to the session. The PCRF then sends a decision to the GGSN. If the PCRF approves the resource reservation, the GGSN sends a PDP Context Activation Response to the UE indicating that the resource reservation has been completed. A similar process takes place at the callee's end.

Steps 23-31: In between steps 22 and 23, there are other events, e.g., 180 Ringing, that take place. These events are omitted to prevent cluttering Figure 3-22. When the callee answers the call, a SIP 200 OK message is sent towards the caller. When the SIP 200 OK reaches the P-CSCF, it will approve the QoS commitment by sending a decision to the GGSN. Upon receiving this message, the GGSN opens the gate, thereby effectively permitting the IP data flow to use the resources reserved previously. Once this is done, the GGSN responds to the PCRF with a report on the status of the session. A similar process takes place at the caller's end. When this entire process is completed, the proper resources on the data path have been reserved and committed to the session.

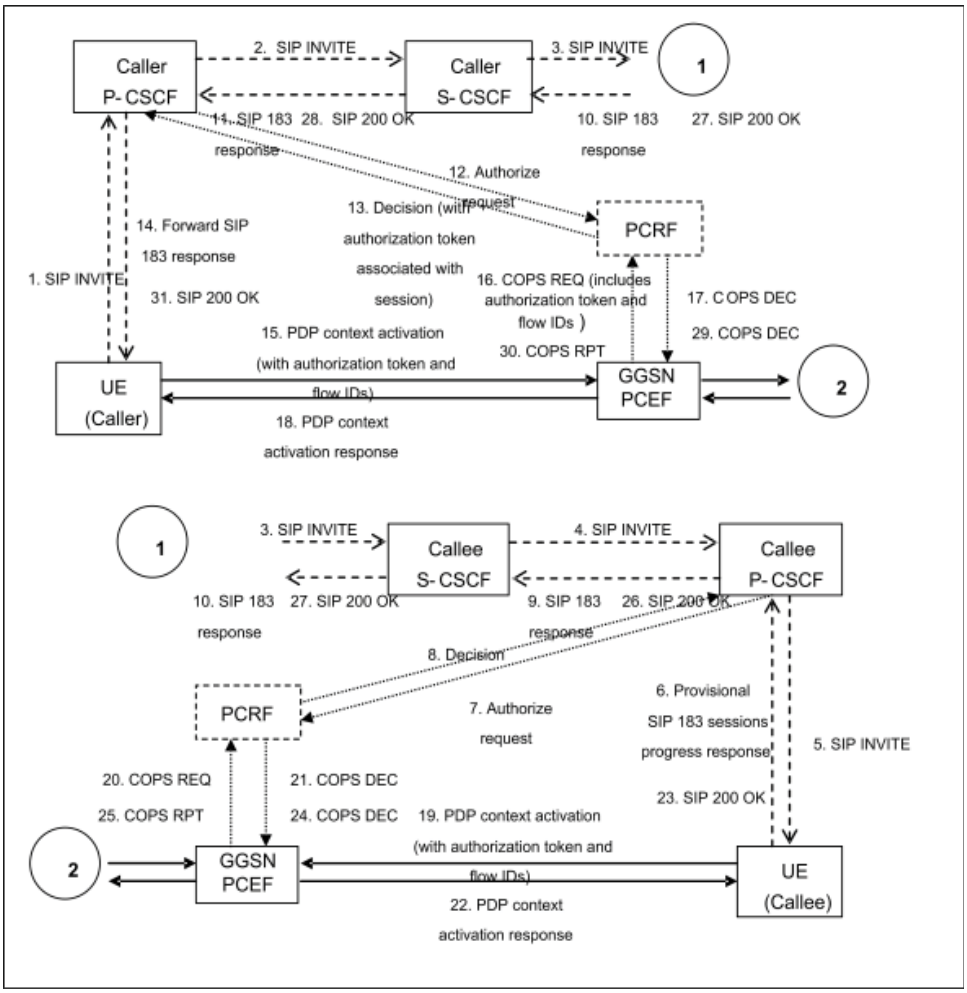


Fig. 11. Session authorization mechanism in a UE-to-UE session establishment process

5. References

Nortel White Paper (2002). Benefits of Quality of Service (QoS) in 3G Wireless Internet, Nortel Networks.

Sudhir Dixit, Yile Guo, Zoe Antoniou (2001). Resource management and quality of service in third-generation wireless network, *IEEE Communication Magazine*, Feb. 2001, pp.125-133.

Sotiris I. Maniatis, Eugenia G. Nikolouzou, & Iakovos S. Venieris (2002). QoS issues in the converged 3G wireless and wired networks, *IEEE Communications Magazine*, Aug. 2002, pp.44-53.

3GPP TS 23.107 (V9.2.0) (2011). Quality of Service(QoS) Concept and architecture (Release 9).

S.Shenker, C.Partridge, R.Guerin (1997), Specification of Guaranteed Quality of Service, *RFC2212*, Sept.1997.

J. Wroclawski (1997), Specification of the Controlled-Load Network Element Service, *RFC2211*, Sept. 1997.

S.Shenker, J.Wroclawski (1999), Network Element Service Specification Template, *RFC2216*, Sept. 1999.

White paper (1999), QoS Protocol & Architecture, *www.qosforum.com*, July, 1999.

R. Braden, D. Clark, S. Shenker, Integrated Services in the Internet Architecture: an Overview, *RFC 1633*, June 1994.

Braden, B., Ed., et. al., Resource Reservation Protocol (RSVP) - Version 1 Functional Specification, *RFC 2205*, September 1997.

S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, An architecture for differentiated services, *RFC 2475*, Dec. 1998.

V. Jacobson, K. Nichols, K. Poduri, An expedited forwarding PHB, *RFC 2598*, June, 1999.

J. Heinanen, F. Baker, Weiss, J. Wroclawski, Assured forwarding PHB group, *RFC 2597*, June 1999.

White paper, Introduction to QoS Policies, *www.qosforum.com*, 1999

Survey on Policy-based networking, *INTAP*.

A. Westerinen, etc. Terminology for Policy-Based Management, <*draft-ietf-policy-terminology-04.txt*>, July, 2001.

B.Moore, E. Ellesson, J. Strassner and A. Westerinen, Policy Core Information Model – Version 1 Specification, *RFC 3060*, IETF, Feb. 2001.

M. Handley, et al., SIP: Session Initiation Protocol, *Internet draft (work in progress),* <*draft-ietf-sip-rfc2543bis-09.txt*>, Feb. 2002

3GPP TS 29.212 (version 8.3.0), Policy and Charging Control Over Gx Reference Point (Rel 8), Dec. 2008.

3GPP TS 23.228 (version 8.3.0), IP Multimedia Subsystem- Stage 2 (Rel 8), June 2008.

3GPP TS 23.203 (version8.3.0),Policy and Charging Control Architecture (Rel. 8), 2008.

W. Marshall, et al., Private SIP Extensions for Media Authorization, *RFC 3313*, Nov. 2001

D. Durham, J. Boyle, R. Cohen, S. Herzog, R. Rajan and A. Sastry, The COPS (Common Open Policy Service) Protocol, *RFC 2748*, IETF, Jan. 2000.

R.Yavatkar, D.Pendarakis, R.Guerin, A Framework for Policy-based Admission Control, *RFC 2753*, IETF, Jan. 2000.

R. Atkinson, Security Architecture for the Internet Protocol, *RFC 2401*, IETF, Aug. 1995

T. Dierks and C. Allen, The TLS Protocol Version 1.0, *RFC 2246*, IETF, Jan. 1999

K. Chan, D. Durham, S. Gai, S. Herzog, K. McCloghrie, F. Reichmeyer, J. Seligson, A. Smith and R. Yavatkar, COPS Usage for Policy Provisioning, *RFC 3084*, Mar. 2001

L-N. Hamer, K. Chan, H. Syed, H. Shieh and R. Zwart, COPS-PR for outsourcing in UMTS: UMTS Go PIB, *draft-hamer-rap-cops-umts-go-00, IETF*, Nov. 2001

B. Moore, L. Rafalow, Y. Ramberg, Y. Snir, J. Strassner, A. Westerinen, R. Chadha, M. Brunner and R. Cohen, Policy Core Information Model Extensions, *draft-ietf-policy-pcim-ext-06*, Nov. 2001

J. Jason, L. Rafalow and E. Vyncke, IPsec Configuration Policy Model, *draft-ietf-ipsp-config-policy-model-03, IETF*, July 2001

Y. Snir, Y. Ramberg, J. Strassner, R. Cohen and B. Moore, Policy QoS Information Model, *draft-ietf-policy-qos-info-model-04, IETF*, Nov. 2001

S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, An Architecture for Differentiated Service, *RFC 2475, IETF*, Dec. 1998

R. Braden, D. Clark and S. Shenker, Integrated Services in the Internet Architecture: an Overview, *RFC 1633, IETF*, June 1994

R. Braden, Ed., L. Zhang, S. Berson, S. Herzog and S. Jamin, Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification, *RFC 2205, IETF*, Sept. 1997

B Moore, D. Durham, J. Strassner, A. Westerinen, W. Weiss and J. Halpern, Information Model for Describing Network Device QoS Datapath Mechanisms, *draft-ietf-policy-qos-device-info-model-06, IETF*, Nov. 2001

M. Wahl, T. Howes and S. Kille, Lightweight Directory Access Protocol (v3), *RFC 2251, IETF*, Dec. 1997

G. Good, The LDAP Data Interchange Format (LDIF) - Technical Specification, *RFC 2849, IETF*, June 2000

Ebata, M. Takihiro, S. Miyake, et al., Interdomain QoS Provisioning and Accounting, *INET 2000*, Yokohama, Japan, July 2000

K. Nichols, V. Jacobson, L. Zhang, A Two-bit Differentiated Services Architecture for the Internet, *RFC 2638*, July 1999.

B. Teitelbaum, P. Chimento, "QBone Bandwidth Broker Architecture", work in progress, http://qbone.internet2.edu/bb/bboutline2.html.

Part 3

Sensor Networks

6

Power Considerations for Sensor Networks

Khadija Stewart[1] and James L. Stewart[2]
[1]*DePauw Universtity*
[2]*Purdue University*
USA

1. Introduction

Wireless sensor networks (WSNs) are networks composed of small, resource-constrained and collaborative devices. WSNs are used in a plethora of domains including environmental and agricultural monitoring, military operations, in the health care field and in building automation. The three main functions of wireless sensor nodes (also called motes) are to sense the environment, perform computations, store intermediate results and communicate with other motes in the network.

This chapter focuses on power considerations for all aspects of wireless sensor networks. It covers software, hardware and networking aspects of the motes. The main limitation of wireless sensor motes is that they operate on battery power. In many WSN applications, the motes are placed in remote areas and deployed for the lifetime of the network. During this time the only power resource the motes have access to is their battery. An example of such a deployment is the Mount St. Helens project developed to study volcanic activities on Mount St. Helens (were volcanic eruptions can occur at any time with very little warning). The sensors were placed on the mountain using helicopters and work at length to continually sense seismic activity and relay information to a data center. For such applications, the battery lifetime is the main factor that dictates the lifetime of the network. It is therefore imperative to develop wireless sensor mote platforms that minimize the power consumption and/or maximize the lifetime of the network as a whole.

Several works in the literature address one or two aspects of the mote's architecture and/or functionality but to the authors' knowledge, no work has combined all said aspects and addressed them as a homogeneous unit. This chapter studies and analyzes each hardware component of the mote's architecture, all the main protocols used in the mote's stack layer, discusses the work that has been done in terms of reducing the power consumption, increasing the battery lifetime and or increasing the lifetime of the entire network as a whole.

The chapter is organized as follows: Section 2 gives an overview of wireless sensor networks, their applications and general architecture. Section 3 focuses on the hardware architecture of the motes (the CPU, communication infrastructure, memory and sensors). Section 4 introduces the layered protocol stack of the sensor motes (application, transport, network, link and physical layers). Section 5 summarizes the chapter and suggest paths forward.

2. Preliminaries

Wireless sensor networks are composed of small, inexpensive devices that are designed to sense some phenomena, perform light computations and communicate with one another. These devices are usually scattered over some area. This technology has seen a wide range of applications ranging from military use to personal security. In the following, we discuss the history of WSNs and some of their most pertinent applications.

Wireless sensor networks evolved form the Smartdust project, which was developed and funded by DARPA in the late 1990s. The Smartdust project was designed to show that "a complete sensor/communication system can be integrated into a cubic millimeter package" (Pister, 2001). The Smartdust motes were engineered to be power efficient. This and other similar projects have led to the explosion of research in the area of wireless Ad Hoc and sensor networks, which was and still is heavily supported by US government agencies including the National Science Foundation. While working on the Smartdust project, the researchers recognized the variety of applications for their work both in the military field and elsewhere.

Some of the applications for the Smartdust projects are virtual keyboard, inventory control, product quality monitoring and smart office spaces among others (Pister, 2001). In the virtual keyboard application, dust motes would be glued into fingernails to sense the orientation and motion of the fingertips and communicate with a computer. This could be used in sign language translations, piano play, etc... In the inventory and quality control applications, a system of communication could be implemented and deployed in all aspects of the production process in order to monitor the location of the product and control and monitor its quality (from temperature, to humidity exposure etc...). In the smart office spaces application, the person's preferred temperature, humidity settings could be directly communicated to the environment they walk into. Some of the military applications that the Smartdust project was developed for include battlefield surveillance, transportation monitoring and missile monitoring.

In the past few year, Wireless Sensor Networks made the transition from the Berkeley research centers to the production arena with the creation of companies, such as Crossbow Technologies (Crossbow Technologies, n.d.) that started manufacturing them. The appeal of Wireless Sensor Networks stems from the fact that you can deploy them and just leave. We discuss in the remainder of this section the main classes of applications for the general WSNs.

WSN applications can be categorized into habitat and environmental monitoring, heath applications, commercial applications, military applications among others.

One of the most prevalent uses of WSNs is in habitat and environmental monitoring. It has been shown that direct human contact with some plant or animal colonies can result in disastrous consequences. For example, (Mainwaring et al., 2002) describe the use of a sensor network to monitor Seabird colonies because of their sensitivity to human disturbance. In fact, a 15 minute visit to the colony could result in up to 20% rate of mortality among eggs. Not only are WSNs useful in monitoring colonies without causing any disturbances but they also represent a more economic method of monitoring for long periods of time.

Another example of the environmental use of WSNs is in forecasting systems. WSNs are now scattered around large areas to forecast pollution, flooding and seismic activity. The Automated Local Evaluation in Real Time (ALERT) was developed in the 70s by the National Weather Service. It has been used by several government and state agencies and international

organizations to provide a real time data collection system that can forecast floods (ALERT, n.d.).

Another use for WSNs is in intelligent building management. In fact, they have been used in HVAC, lighting, climate control, fire protection, energy monitoring and security applications among others. In Canada for example, the National Research Council launched a three-year project to develop wireless sensor networks to do just that. The project started in 2008 and is anticipated to continue through 2011.

A very important application of WSNs is in the healthcare field. WSNs can be used to provide continuous, remote, inexpensive, instantaneous and non-invasive monitoring of a patient's vital signs. This technology can be used to allow the elderly to remain in their own residences but still be able to continuously check their vitals.

All these WSN applications consist of deploying the network for an extended period of time on a single battery charge. It is therefore imperative that the motes be power efficient and that the lifetime of the network as a whole be as long as possible.

3. The WSN hardware architecture

The hardware of wireless sensor motes consists of sensors (analog and/or digital), a microcontroller, also referred to as a microprocessor or Central Processing Unit(CPU), memory, RF communication module (transceiver) and battery. The design of each component of a WSN mote should take into consideration the power metrics (power consumption and voltage requirements) of the component. Additionally, the integration/interface of all the components as a whole should be studied for power consumption (having analog sensors means that an ADC component in the CPU should be required to convert the sensor readings to a digital format etcÉ)

To reduce power consumption, several works suggest the introduction of sleep and wake up cycles for the motes. Other schemes suggest a better integration of the functionality of hardware components (using cross-layer principles). Another consideration in the design of the CPU is the clock component. Several applications of WSNs require some level of time synchronization. Clock choices and designs affect the amount of drift that a sensor mote's clock can experience requiring more or less time synchronization operations when the mote is deployed (Akyildiz et al., 2002).

3.1 Sensors

Of the five main units, the sensing unit is the most application specific. Meaning the type of sensor used will depend on the application. For instance, wireless sensors used for structural health monitoring may consist of materials apt for monitoring strain, acceleration (accelerometer) and linear and angular displacement. Other application specific sensors may measure, vehicular movement, soil consistency, blood alcohol levels, humidity, noise levels and so on. These sensors should then report some signal indicative of their acquisition. Temperature (thermo-coupler outputting voltage or thermistor outputting resistance), force or pressure (piezoelectric outputting voltage, strain gauge outputting resistance), position (linear variable differential transformers (LVDT) outputting alternating current) or light intensity (photodiode outputting current) all need to report information regarding their surroundings to a processing unit (Wilson, 2004).

The development of an efficient method for acquiring and converting conventional energy from the sensors such as solar and wind has seen an exponential growth over the last few years. The sensing development has been referred to as energy harvesting. One factor contributing to the enjoyment of such an increase has been the threat of rapid decreases projected in our global and national energy reserves based on utilization rates and trends. Such a premonition has spurred funding for research in various fields including materials or more specifically metamaterials.

Metamaterials have been defined by most associated scientist as materials made by man which exhibit non-natural properties and characteristics, particularly EM or electromagnetic properties not known to exist with any other materials found in nature. Regarding electromagnetism, metamaterials which exhibit propagating electromagnetic waves (both the permittivity and permeability are negative as seen in Figure 1 have seen much attention in recent years as well as when both permittivity and permeability are very close to 1.

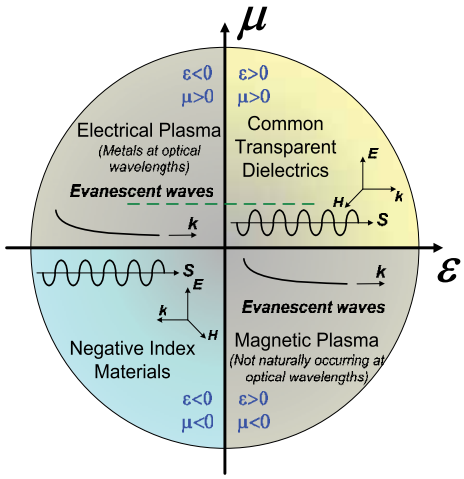


Fig. 1. The parameter space for ϵ and μ. The two axes correspond to the real parts of permittivity and permeability, respectively. The dashed green line represents non-magnetic materials with $\mu = 1$ Cai &shaleav (2010).

The reaction or response of a material (as in a sensor for WSNs) to external fields is largely determined by only the two material parameters ϵ and μ, permittivity and permeability respectively. As shown in Figure 2, the real part of permittivity ϵ_r is plotted to the horizontal axis of the parameter space, while the vertical axis corresponds to the real part of permeability μ_r. A negative value of ϵ (μ) indicates that the direction of the electric (magnetic) field induced inside the material is in the opposite direction to the inbound incident field. Noble metals at optical frequencies, for example, are materials with negative ϵ, and negative μ and can be found in ferromagnetic media near a resonance. Waves can not propagate in material in the second and fourth quadrants, where one of the two parameters is negative and the index of refraction becomes purely imaginary. In the domain of optics, all conventional materials are confined to an extremely narrow zone around a horizontal line at $\mu = 1$ in the space, as represented by the dashed line in Figure 2.

Scores of such materials are designed to manipulate EM waves, many passively, by creating an alternate propagation path. Metamaterials have been designed to redirect, not absorb or

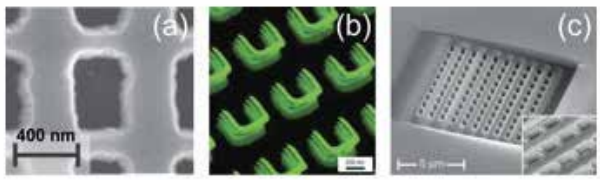


Fig. 2. Examples of a 3D optical metamaterials fabricated in a layer-by-layer manner. (a) A near-infrared NIM (Negative Index Material) with three functional layers made by EBL (Electron Beam Lithography); (b) Four layers of SRRs (Split Ring Resonators) based on EBL with patterning-and-flattening approach; (c) A NIM wedge exhibiting negative refraction for visible light made by an advanced FIB technique Cai &shaleav (2010).

reflect but to divert the energy through a desired path. It is no wonder as to the attention metamaterials have seen for energy harvesting. Research is currently being conducted to develop sensors (both photon and electron based) that extract atmospheric energy regardless of the incident angle such that no energy is reflected back out of the sensor rather, it's reflected down toward the detector. This will lead to the creation of ultra-efficient sensors for wireless networks, see (Narimanov & Kildishev, 2009; Shalaev et al., 2005) for more information on Metamaterials.

3.2 Microcontroller

The component responsible for doing the bulk of the switching and decision making for the WSN at the remote site is the processor or microcontroller. When selecting the processor for specific WSN applications, the engineer must make many considerations. These considerations include, but are not limited to, cost, power requirements, physical size, weight and speed, some of which are elaborated upon below.

Depending on the microcontroller, the power requirement could range from .25 mA to 2.5 mA per MHz for either 8 or 16 bit processors. Again, the application will determine if a processor consuming relatively high amounts of energy is acceptable or if .25 mA per MHz is needed. A common misconception is that by putting the processor in "sleep" mode, the sensor utilizes less power thus is more efficient. It has been shown that this is not always true as while in "sleep mode", sensors still maintain synchronization and memory functionalities necessary to perform expeditiously upon awakening (Hu & Cao, 2010).

In fact a more prudent approach to saving energy would include completely shutting the processor off, entirely, and ensuring the sensor can rapidly recover from a "dead" state or at the very least rapidly jump from "sleep" mode to "awake" mode. As the processor needs to synchronize native clocks and stabilize, the transition time or delay can be as long as 10 ms which is a relative eternity. Another parallel approach involves varying the speed depending on the time allotted for a specific task.

In other words, only using the minimum power required for a task at a given time by dynamically ramping up or down the power accordingly versus drawing full power for all "awake" states. This approach may benefit from an algorithm in which the speed is a function of the power. If the required task and itÕs effort expended is known before the task is given, an absolute "finish time" can be maintained without necessarily completing the task as fast

a possible rather as fast as necessary. Researchers from the University of California, Irvine (Irani et al., 2007) developed an algorithm for optimizing power consumption by varying speed below:

$$g(z,z') = \frac{\sum_j suchthat[r_j, dj] \subseteq [z,z']R_j}{l(z,z')} \quad (1)$$

where g(z,z') defines the intensity of the interval [z,z'], l[z,z'] defines the length of the interval, R_j is the required work needed to complete the job and d_j denotes the deadline for job j. This would allow energy and speed to be spent where it's needed most creating a dynamic fluid speed variance throughout the CPU for maximum overall efficiency. One might say, 'losing a battle here and there but winning the war'.

3.3 Memory

Memory is a crucial factor in WSNs. Particularly non-volatile memory. Non- volatile memory is defined as various forms of solid state memory which doesn't need to be refreshed or powered to maintain it's information. Examples include flash, electrically erasable programmable read-only memory (PROM) read only memory (ROM), optical discs and magnetic disks (Postolache et al., 2010).

The memory component is the means at which the WSN stores the data it acquires. The speed requirement of the memory unit depends of the nature of the WSN and its intended functionality. A rather fast memory unit may be required for certain military applications where the data acquisition speed from the memory may dictate whether or not a target is detected in time for acquisition and lock. On the other hand, a relatively slow memory unit may be acceptable for soil monitoring WSN utilized by farmers. In either case the security and reliability of the memory unit is important and both require additional power demands on the WSN. To this end, researchers have been developing ways to more efficiently processing and storing the acquired data including virtual memory protocols. Virtual memory has been shown to reduce compile-time optimizations regardless of the limitations in memory on site. One approach which generates a memory layout optimizes to the memory access patterns and attributes for a given WSN. In other words, the protocol optimizes the memory map based on the application, effectively reducing overhead (Lachenmann et al., 2007).

3.4 Transceiver module

All WSN motes will possess a transceiver or TR modules as they allow the motes to communicate in WSNs. They present the capability to transmit and receive data packets, information or signals in a relatively small package. One of the main factors which allows for such a diminutive size lies in the RF architecture. Because the TR modules transmit and receive in the same RF component there is no need for a separate architecture for each transmission or reception. Thus the isolation of incident energy to transmitted energy must be great to ensure against destructive cross modulation, unwanted dispersion and various other resultant noise, all of which would inherently degrade the efficiency of the WSN either directly or indirectly. Signal loss is of particular concern in the input/output portion of the TR module and precautions must be taken to ensure signal degradation is tolerable from a minimum threshold point of view.

Within the TR package, a typical TR module will consist of and follow this RF path for transmission: a common attenuator for signal suppression, a common phase shifter (depending on the application. For example, phase shifter could be used to shape the transmission pattern or radiation pattern leaving a WSN (also known as beam-steering), a driver and a high power amplifier (HPA) to boost the signal amplitude for propagation from the aperture or antenna for transmission. When receiving a signal within the TR module frequency range, which varies per application, the signal passes through a limiting filter and low noise amplifier (LNA) before coursing through a common attenuator to suppress the signal's magnitude and possibly a common phase shifter (depending on the application. For example the phase shifter can be utilized as a directional finder or filter for incident signal in a WSN). Outbound and incident signals are typically discerned by a circulator at the output/input of the module. The attenuator and phase shifters are termed "common" due to the fact that these components are used for both reception and transmission. In the following, we elaborate on a few of the key components of the TR module from Figure 3.

The attenuator is implemented to ensure the unwanted side-lobes are suppressed, sufficiently reducing the noise in the system. It also keeps the amplifiers down stream from prematurely reaching saturation and causing unwanted non-linearities. Typically this is done only for the receiver as during transmissions, it is usually desirable to propagate as much energy from the antenna as possible. Since the attenuator basically performs the exact opposite function of the amplifier, they are typically not conjoined in series unless, in some cases, it's needed for filtering purposes. Note that all the components within the TR module are frequency matched meaning they are optimized for specific frequency ranges. Due to this inherent characteristic, attenuators can be used to suppress frequency bands without distorting the fundamental waveform. This is important for the energy efficiency of the system as the modulator can maintain relative simplicity without the need to effectively recreate a waveform which would subsequently cost more power.

The phase shifter allows multiple RF signals to be controlled by way of an external stimulation such that the output of the phase shifter is of the desired phase without effecting the frequency. The phase shifter may or may not be present in the TR module. It depends on whether or not the WSN calls for a beam-forming or shaping capability which can aid in power efficiency if multiple sensors are synchronized in receive and or transmission mode for power/amplitude coupling. The amplifiers (driver and high power amplifiers) boost the signal for transmission from the antenna. The level of amplification needed depends on the efficiency of the system, particularly the aperture or antenna. A poorly matched antenna or one which has a high Voltage Standing Wave Ratio (VSWR) will demand a higher amplitude or stronger signal to propagate to a given target.

The application and placement availability of WSN will greatly affect which antenna is more suitable and efficient. Most WSN antennas are omni-directional fundamentally but are shaped by various ground effects. This crucial aspect of antenna propagation has prompted many researchers to develop accurate prediction models specifically for WSNs.

3.5 Power source

Considering that many WSNs rely on portable energy or power sources to power sensors, the capacity and efficiency of both the power source and the WSN is crucial in the overall effectiveness of the WSN. For most of the WSN applications, when the power source drains,

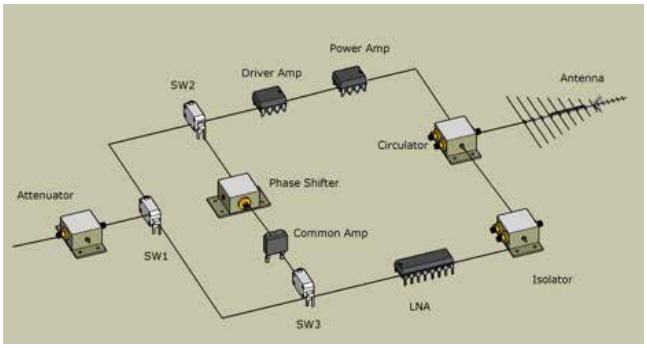


Fig. 3. Transceiver

the WSN is inoperable. For many applications various protocols for maximizing the lifetime of the WSN are adequate while many other applications require WSNs to remain in remote areas for several months or years without opportunities for manual power replenishment. Many research centers have developed models to efficiently harvest energy for power as for sensing previously mentioned. A WSN which can obtain its power requirements from its surrounding environment essentially has an infinite lifetime. Various approaches from mechanical vibrational energy harvesting to photon collection schemes are being considered in an effort to self-generate power needs.

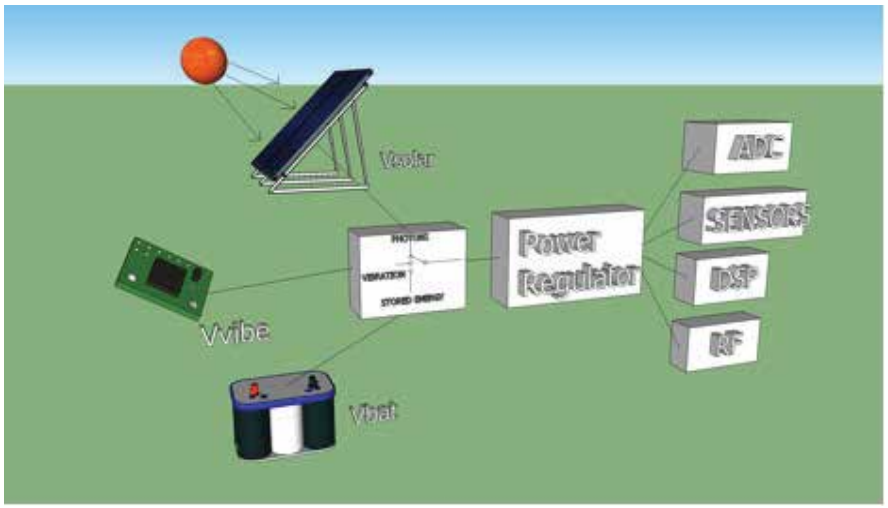


Fig. 4. A low power wireless sensor node system powered from energy scavengers or harvesters and a battery. Guilar et al. (2006).

Figure 4 is a schematic of a low power WSN system that uses energy scavengers. In Figure 4, the energy sources are labeled Vsolar, Vvibe and Vbat for the solar, mechanical vibration and battery, respectively. A mutiplexer switches between the unregulated energy sources. ADC denotes the Analog to digital converter, DSP denotes the Digital signal processors and RF denotes Radio frequency.

4. The WSN layered protocol stack

The WSN layered protocol stack consists of the Application layer, the Transport layer, the Data link layer and the Physical layer. This section will cover the role of each layer and study its power consumption. The section will survey the current literature and analyze it with respect to power consumption.

4.1 Application layer

The application layer is in charge of collecting and processing sensor readings (including the use of data aggregation), performing time synchronization, implementing a security protocol (as needed) etc... Each one of these tasks uses one or more hardware modules and each task results in power being consumed.

4.1.1 Information fusion

Traditionally, sensor motes were designed to perform very little to no processing. They would sense the environment and send the sensing data to the base station for processing. This resulted in large amounts of packets being sent from the motes to the base station. In addition, in several sensor network applications, the motes are exposed to conditions (such as very high/low temperatures, pressure and noise) that might sabotage their measurements. It was then proposed to use information fusion (also referred to as data aggregation) techniques at the motes in order to decrease the network traffic, save energy, remove outlier data, make predictions about future measurements and in general obtain better information quality by combining data from multiple sources. Data aggregation requires some amount of processing to be carried out at the motes. Data fusion can be used at different layers of the WSN protocol stack. For example, it can be used at the application layer to process sensor readings as well as at the network layer to consolidate routing information. In the following, we survey and analyze the work that has been done on data aggregation and information fusion.

Information fusion can be categorized into three classes. Complementary, redundant, and cooperative. This classification depends on the particular application and the relationship between the motes that gather the data. In the case of complementary information, sources gather different types of data and information fusion is applied to obtain a more complete picture from data. In the case of redundant information, one or more sources gather the same type of data and information fusion is used to discard the outlier measurements and filter the data for accuracy, reliability and confidence. In cooperative information fusion, two sources gather information that is fused to produce information that better represents the reality. Information fusion is performed for different purposes. In the following, we present a classification of data fusion algorithms based on the purpose of the information fusion.

Information fusion techniques could be either centralized or distributed. Centralized techniques have a single point that controls the fusion process but are simple to implement. However, all the sensor motes send their data to the central point, which overwhelms the central data point and floods the network with messages. Distributed techniques on the other hand are more complex to implement but are more energy efficient because the information is exchanged locally, which reduces the number of messages exchanged in the network. Several methods have been proposed for information fusion including inference, estimation, aggregation, and compression.

Inference methods consist of making a decision based on previous knowledge. Protocols that have been proposed for inference include Bayesian inference algorithms (Coue et al., 2002; Tsymbal et al., 2003), Dempster-Shafer inference algorithms (Dempster, 1968; Shafer, 1976) and fussy logic algorithms (Gupta et al., 2005) among others.

Estimation methods use probabilistic theory to estimate a state based on a sequence of measurements. Estimation algorithms include the maximum likelihood algorithm (Xiao et al., 2005), least square, Kalman filter and particle filter (Kalman, 1960).

Data aggregation methods are used to overcome implosion and overlap and compression is used to reduce the amount of data by exploiting spatial correlation among the motes. Techniques for compression include distributed source coding (Xiong et al., 2004) and coding by ordering (Petriovic et al., 2003).

The implementation of these algorithms comes at a cost involving hardware complexity, CPU time and energy.

4.1.2 Time synchronization

Time synchronization consists of synchronizing the local clocks of all the members of a distributed network. In WSNs, it consists of synchronizing the clocks of all the motes in the network. Time synchronization is essential for all networked systems and is a requirement in most WSN applications and protocols. Example applications include environmental monitoring and target tracking among others. In these applications, the order of events is usually important. For example, in target tracking, sensors need to continuously report the location of a moving target, which could be time sensitive. Example protocols that require time synchronization are some MAC protocols (Demirkol et al., 2006) (such as the ones based on TDMA, where each node is assigned a time slot) in addition to several routing and security protocols.

In this section, we review the time synchronization algorithms proposed in the literature and analyze their power saving properties. In addition to being energy efficient, time synchronizations schemes for WSN need to be accurate and scalable.

When a packet is sent from node A to node B, node A can append a time stamp to the packet. Node B can then extract the time stamp from the packet, add the time it took the packet to travel from node A to node B (transmission time) in order to estimate its local clock's drift from node A. The packet delay consists of send time, access time, propagation time and receive time. Send time is the time interval between when the node issues the send command until the node is ready to send the packet.

The medium access times is the duration from when the node is ready to send until the time when the transmission starts. This is the step that makes time synchronization such a difficult problem. It is not possible to accurately estimate this time. The propagation time is the time it takes the packet to reach to the destination, and the receive time is the time it takes to receive the frame.

The Network Time Protocol (NTP) described in (NTP, n.d.), is the protocol that synchronizes the clocks in wired networked systems by estimating the roundtrip time of packets. It is the standard used on the Internet. NTP maintains a universal time (UTC) across the network. NTP is not suitable for WSNs because of its centralized nature and prohibitive cost. In fact

in NTP, clients synchronize their clocks to the server and servers are synchronized to using external time sources (using a GPS). NTP is not suitable for WSNs for a number of reasons. First, NTP is centralized. 2. In WSNs, it is impossible to accurately estimate the roundtrip time. 3. GPS is too expensive to use or is not an option for most WSN applications (for example, indoor applications will not have access to GPS signal).

Most time synchronization protocols are sender to receiver. The sender time stamps a packet and the receiver extract the time stamp and tries to extrapolate its clock drift compared to the sender (Romer, 2001), (Ganeriwal et al., 2003). However, the Reference Broadcast Scheme (RBS) (Elson et al., 2002) is different. It is a receiver-to-receiver synchronization protocol. In RBS, a sender broadcasts a beacon without any time information. The receivers then exchange Acknowledgement messages with the time they received the beacon. Receivers can then extrapolate their own clock drift relative to each other. RBS works with two receivers and is easily extended to more than two receivers. In addition, increasing the number of broadcasts increases the accuracy of the scheme. Note that in RBS, the uncertainty of access time is removed (since the sender is removed from the drift calculations) and since the propagation time is assumed to be negligible in WSNs, the only uncertainty factor and potential error margin in this protocol is the receive time.

The Timing sync Protocol for Sensor Networks (TPSN) (Romer, 2001), (Ganeriwal et al., 2003) is a sender-receiver protocol. In TPSN, the sender sends a packet to a receiver, which uses the TPSN equation to extract its local clock drift compared to the sender. TPSN then uses a tree hierarchy to propagate the synchronization, it categorizes the nodes in the network into levels during the discovery phase. During the synchronization phase, the root node (level 0) synchronizes all level 1 nodes. After this first phase of synchronization, level 1 nodes synchronize level 2 nodes and so on, until synchronization has been propagated through the entire network. TPSN achieves better accuracy than RBS when using MAC layer time stamps because RBS is limited by the transmission range and would require more beacons in order to perform synchronization.

In (Greunen & Rabaey, 2003), the authors claim that most sensor network applications do not require very precise synchronization. In fact, they claim that most applications only require synchronization in the order of a fraction of a second. The authors therefore propose a different approach where the required accuracy is taken as a constraint and then a synchronization algorithm with minimal complexity is devised so that the requested accuracy can be achieved. In this work, the synchronization is propagated in a centralized manner where a spanning tree is created and synchronization is conducted along the edges of the tree.

Centralized approaches to time synchronization are not energy efficient and often result in depleting the energy reserves of the root node. The authors in (Maroti et al., 2004) propose the Flooding Time Synchronization Protocol (FTSP). FTSP uses periodic flooding of synchronization messages. This approach makes the algorithm de-centralized, scalable and topology independent. In FTSP, the synchronization root is elected dynamically and re-elected periodically. The root is responsible for keeping the global time of the network. In this work, the nodes form a dynamic mesh like structure to propagate the time synchronization throughout the network (unlike TPSN). This work saves on the energy required to create an initial spanning tree (as in TPSN) and is therefore more energy efficient than TPSN. In addition, this protocol is not topology dependent and can perform synchronization even when

the topology of the network changes. However, the synchronization error in FTSP can grow exponentially with the size of the network (Lenzen et al., 2009).

Similar to FTP, the Novel Algorithm for Time Synchronization (NATS) is a decentralized time synchronization protocol. Unlike FTSP, NATS is a receiver-sender protocol because the receiver requests synchronization from the sender. This reduces the amount of messages exchanged for the purpose of time synchronization and therefore, reduces the amount of energy consumed during synchronization. NATS was designed at DePauw University by Peter Terlep [1] , Steven Klaback [2] and Khadija Stewart. NATS does not need to meet any specific topology prerequisites, it can adjust to topology changes. It accomplishes the following: 1) it does not need a third party device that is within radio communication range of all motes, 2) it does not need any one mote to be within range of all motes, 3) it is scalable, 4) it allows for deep sleep between synchronization activities, 5) it handles receiver-side medium access control (MAC) buffer latency uncertainty, 6) it addresses the inability to acquire a real-time sender-side MAC timestamp, 7) and it uses a distributed energy efficient algorithm for multi-hop synchronization.

Pair-wise synchronization in NATS starts when the root node receives a sync request. The root then sends two consecutive packets to the requesting node, each containing a timestamp at the MAC layer. The receiving node uses these two packets along with its receive time stamp to extrapolate the propagation and channel access times. It uses that information to estimate its clock drift from the root node. Time synchronization is then propagated throughout the network in a distributed manner, similar to FTSP, by having each synched node act as a potential root node for synchronization. Experimental results show that NATS provides better synchronization accuracy than TPSN. In fact, using the Sun Spots platform, the Mean Sync Error for NATS was 1.74ms versus 2.63ms for TPSN.

The Gradient Time Synchronization Protocol (GTSP) is completely distributed (Sommer & Wattenhofer, 2009), where the nodes periodically broadcast synchronization beacons to their neighbors and agree on a common clock. It is proven that after multiple beacon exchanges, the clock of the nodes converges to a common value. This algorithm is completely distributed and nodes only exchange beacons locally. GTSP is proven to achieve better time synchronization accuracy as compared to tree-based methods.

In PulseSync (Lenzen et al., 2009), the root node floods a "pulse" through the network in a breadth-first search tree manner. The nodes receiving the pulse then compensate for the drift relative to the root node. The authors note that the flooding of the pulse needs to be scheduled in order to avoid collisions. This protocol is proven to be accurate when used in sensor network applications where the topology does not change. In fact, it is proven to outperform FTSP by a factor of 5 on mid-size networks.

The authors in (Li et al., 2011), propose a new direction in time synchronization where the Radio Data System (RDS) of FM radios is used to synchronize the nodes' clocks. In this work, each node is equipped with an FM receiver and programmed to receive the same RDS signal. The mote's clock then uses a calibration component to calibrate itself to the RDS clock. The drawbacks of this method stem from the fact that the FM interface is not power efficient and

[1] Peter Terlep is currenty a Ph.D student at Michigan University
[2] Steven Klaback is currently with Digital Knowledge

that not all WSN applications can have access to FM signals especially for the applications deployed in remote areas.

In summary, in order for a time synchronization protocol to be appropriate for a wide range of WSN application, it needs to accurately compute the clock drifts, be distributed, scalable, adapt to any topology and be able to propagate the synchronization instantaneously and without flooding the network.

4.2 Transport layer

The transport layer is mainly used to communicate with external networks (such as the Internet) and is therefore rarely implemented in sensor motes.

4.3 Network layer

The network layer is in charge of all routing functions. Routing is the function that is used the most in multi-hop WSNs. It is the routing algorithm that allows nodes that are more than a hop away to communicate with each other and form a connected network. Because routing is used extensively in most WSN applications, it is the function that should be the most power efficient. A variety of routing protocols have been proposed in the literature, some of which are designed to be 'power aware' and use the battery level or the network lifetime as a routing constraint. This Section reviews these works and studies the effect of clustering on power consumption.

Initially, research on routing algorithms focused on Mobile Ad hoc NETworks (MANETs). In these networks, the nodes were designed to be highly mobile, which resulted in the development of on-demand routing algorithms. These algorithms use flooding to compute routes (see the Reliable Ad-Hoc On-Demand Distance Vector Routing Protocol (RAODV) (Khurana et al., 2006), and the Ad hoc On-demand Multipath Distance Vector (Marina & Das, 2001) among others). The traditional flooding method consists of every node broadcasting the data to all its neighbors, the neighbors broadcasting the data to their neighbors etc... Ultimately, the sink will overhear the data. Flooding-based protocols suffer from several inefficiencies including overwhelming the network with unnecessary transmissions, excessive energy consumption, implosion, overlap, among others see (Heinzelman et al., 1999). Routing in MANETs is a tedious problem because of their dynamic nature. Adding power efficiency to the equation renders the problem even more tedious.

In (Mleki et al., 2002), the authors propose a reactive Power-aware Source Routing (PSR) protocol for MANETs. This protocol was based on the Dynamic Source Routing protocol (RFC4728, n.d.). PSR computes the cost of routes while taking into consideration both transmission power and remaining battery power. In PSR, the source broadcasts a message and intermediate nodes compute the path cost and add it to the header of the broadcast message. The destination then adds the least cost path to the reply and sends it back to the source. This solution fits the needs of MANETs but because of its broadcast nature, it is not suitable for the more resource constrained sensor networks. Since most sensor network applications require static sensors and are more resource constrained than MANETs, the routing solutions that were developed for MANETs are not suitable for the low power sensor networks. As a result, the Routing Over Low power and Lossy networks (ROLL) group was created as part of IETF in 2008 (Watteyne & Richichi, 2010) to help develop a standardized routing solution for sensor networks.

In (Watteyne & Richichi, 2010), the authors define a set of criteria that routing protocols must possess for routing in low-power and lossy networks. These criteria consist of satisfactory performance in: 1. Routing state. 2. Loss response. 3. Control cost. 4. Link cost. 5. Node cost. The authors then conclude that none of the mature IETF protocols, that were developed for MANETs, fulfill those requirements. The protocols examined in this work are: OSPF (RFC2328, n.d.), IS-IS (RFC1142, n.d.), OLSR (RFC3626, n.d.), OLSRv2 (draft-ietf-manet-olsrv2-12, n.d.), TBRPF (RFC3684, n.d.), RIP (RFC2453, n.d.), AODV (RFC3561, n.d.), DYMO (draft-ietf-manet-dymo-mib-04, n.d.), DSR (RFC4728, n.d.), IPv6 Neighbor Discovery (RFC4861, n.d.) and MANET-NHDP (draft-ietf-manet-nhdp-15, n.d.). In (Watteyne & Richichi, 2010), the authors suggest that a new protocol specification document needs to be created for routing in low-power and lossy networks. The discussion in (Watteyne & Richichi, 2010) was limited to mature and well documented IETF protocols, in the remaining of this section, we examine "energy aware" routing protocols designed for wireless sensor networks that have not been included in this review.

Routing algorithms with energy considerations aim to either minimize the energy consumption of the networks as a whole or increase the lifetime of the network. Protocols that attempt to minimize the energy consumption of the network usually compute and use the shortest paths in the network. As a consequence, a few select motes are usually overused and their energy reserve is depleted earlier than the rest of the motes. This could result in the network becoming partitioned and could therefore end its useful lifetime prematurely.

Most applications of WSNs are deployed in remote areas and are scheduled to monitor the area for long periods of time. In this case, extending the useful lifetime of the network is of at most importance. The concept of 'lifetime of the network' is difficult to define in WSNs (Dietrick & Dressler, 2009). For practical purposes, we define the useful lifetime of the network as: 'The total amount of time that the network is able to do useful work'. If for example the purpose of the network is to record sensor readings from ten different areas for as long as possible, the useful (operational or functional) lifetime of the network will be the total amount of time that at least one sensor is functional in each of the ten different areas and that there exists a path between each of those senors to the sink, i.e., those sensor motes are able to relay their readings to the sink. The useful lifetime of the network is therefore application specific and a uniform definition may not apply to all types of WSN applications.

The shortcomings of the broadcast-based protocols have led to the design of data-centric routing mechanisms. One of the earliest works on this type of protocols is SPINS (Heinzelman et al., 1999) where the data is named using high-level descriptors (meta-data). In this case, sensors exchange meta-data. The protocol relies on three types of messages: 1. ADV message, which is used to advertise particular meta-data, 2. REQ message used to request specific data, and 2. DATA message used to deliver the actual data. Spins achieves significant energy savings over traditional broadcast-based protocols (a factor of 3.5) and reduces the data redundancy in half. However, Spins does not guarantee the delivery of data to the requesting node, which makes this protocol unpractical for several applications of WSNs (Akkaya & Younis, 2003).

In data-centric routing algorithms, regions of sensors are queried to send their sensed readings to the sink. Because of the redundancy in sensors in each region, the data needs to be aggregated before it is forwarded to the sink. Several algorithms have been proposed to perform data aggregation to disregard the redundant information. Sensor Protocols for

Information via Negotiation (SPIN) (Kulik, 1999) was the first work to suggest eliminating redundant information to save energy. Later, a series of protocols that use directed diffusion were proposed (Intanagonwiwat et al., 2000), (Braginsky & Estin, 2002), (Schurgers & Srivastava, 2001), (Chu et al., 2002).

An important step in routing in wireless sensor networks was the creation of routing algorithms based on directed diffusion, the first introduction is described in (Intanagonwiwat et al., 2000). In directed diffusion, a node sends a query for some particular data (data here is identified using an attribute-value pair). As a result, data matching the query description is "drawn" towards the querying node. The data can be aggregated by intermediate nodes and all the communication is only neighbor-to-neighbor. These types of algorithms achieve significant energy savings over the traditional broadcast-based algorithms. Despite the energy saving properties of the directed diffusion algorithms, they are not suitable for all sensor network applications. Some sensor network applications require continuous data flow from the sensors to the sink, as a consequence, query based algorithms will not be suitable for such applications since the sink would need to continuously query each sensor for data (Akkaya & Younis, 2003).

An alternate way of relaying information in WSNs, other than flooding, is gossiping (Kyasanur et al., 2006). In gossiping, the source node selects a random neighbor and forwards the data to them. The process continues until the destination is reached or a maximum number of hops is achieved. Similar to flooding protocols, gossiping protocols also waste energy by sending messages by sensors that cover overlapping areas. In addition, gossiping algorithms can suffer from excessive delays because the next hop node is selected randomly.

An improvement to the traditional gossiping protocols is the location-based protocols. In these protocols, location information is used to direct the routing in order to reduce the number of transmissions and therefore save energy. One such protocol is SPEED (He et al., 2002). This protocol uses a combination of feedback control and non-deterministic geographic forwarding to provide real-time unicast, area-multicast and real-time area-anycast.

In (Li et al., 2001), the authors propose an energy saving routing scheme called the adaptive max-minzPmin scheme. This routing algorithm selects a route that maximizes the minimum residual energy as long as it consumes no more than zPmin energy (Pmin energy is he amount of energy consumed by the minimum-energy route). This algorithm also computes the minimum node lifetime of the network and adjusts its routing criterion accordingly. While this method is hard to implement (keeping track of the lifetime of the nodes in a central location), it is more practical for ad hoc networks than it is for sensor networks.

Another family of protocols is the hierarchical routing protocols. The main purpose of creating a hierarchy within a sensor network is to achieve scalability, i.e., the network performance should decrease slowly in response to an increase in the network size. The main form of hierarchical routing in WSNs is clustering, which consists of organizing the nodes into clusters where each cluster has a cluster head. The cluster head is then in charge of performing data aggregation or forward the packets on to the next hop. This leads to a smaller amount of data being transmitted to the sink, which intrinsically saves energy.

One of the first clustering protocols, LEACH is described in (Heinzelman et al., 2000). LEACH randomly rotates the head cluster in order to balance the energy consumption amongst the nodes in the cluster and uses data fusion in order to reduce the amount of data sent to the sink.

As a result, LEACH achieves significant energy savings compared to conventional routing protocols. Several other hierarchical protocols have been proposed in the literature who build up on LEACH such as TL-LEACH (Loscri et al., 2005) which proposes a two-level hierarchy to LEACH, EECS (Ye et al., 2006) where nodes compete for the position of cluster head, HEED (Yonis & Fahmy, 2004) where cluster heads are selected based on the distance between nodes, among others.

In (Iwanicki & Steen, 2009), the authors develop a framework to test the various hierarchical routing protocols proposed for WSNs. The authors state that hierarchical routing is a promising solution for the resource constrained WSNs and caution that the theoretical results presented in most hierarchical work can be very different from the results obtained using a more realistic framework. The proposed framework dismisses the idea of rotating the cluster head to save energy because this change complicates route computation by changing the routing addresses. The authors conclude that there is no one optimal hierarchical routing protocol for all WSN applications, rather protocols are application and requirement dependent.

In conclusion, there still exists the need to develop a low-frills, low-power, manageable and adaptable protocol for routing in the resource constrained sensor networks. The ROLL working group is still working on a requirement specification document. They may in fact, not be able to propose a single protocol for all or most WSN applications and could end up proposing or extending more than one protocol.

4.4 Medium access control layer

The main duties of sensor motes are communication, sensing and computing. Amongst these three tasks, communication consumes the most energy. It is therefore imperative to make sure that the communication task is as efficient as possible in order to prolong the energy lifetime of the motes. It is the data link layer that is responsible for establishing communication links between the motes, allowing the motes to share the wireless medium fairly and detecting/correcting transmission errors. Power considerations at the data link layer involve studying the hardware of the communication module (see Section 3) , the implementation of protocols such as the power management protocol and manipulating the power level of the transceiver.

The most energy waste occurs when a mote receives multiple frames at the same time. In this case all the frames that collide need to be discarded which results in wasted transmissions and receptions and increased latency. Other causes of energy waste are control packet overhead, overhearing unnecessary traffic and the long idle time in WSNs. In fact, in WSNs, idle listening consumes more than half the amount of energy required for reception (Ye et al., 2004). The Medium Access Control (MAC) layer is the sublayer of the data link layer that is responsible for handling the contention over the medium (in this case, the wireless medium). The main media access protocols used in wireless networks are Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), Carrier Sense Multiple Access (CSMA), Request To Send/Clear To Send (RTS/CTS) protocols, and the IEEE 802.11 protocol. The purpose of these schemes is to avoid channel contention. In the following, we review the most relevant MAC protocols that are proposed for use with wireless sensor networks. The channel contention scheme in these protocols is based on the above described contention

prevention mechanisms. In the rest of this Section, we study the main MAC layer protocols that are proposed in the literature and analyze their power-saving properties.

In this work, we consider energy efficiency to be the most important attribute in a MAC protocol. Other important attributes for a MAC protocol consist of providing fair and efficient access to the medium, scalability and adaptability to change.

Most ad hoc network and WSN applications require the network to be deployed for an extended period of time. During their deployment, the motes are programmed to sense the environment and relay sensor readings to the sink. Several MAC protocols have been proposed for these applications where the motes are periodically scheduled to be in a power-saving state (a sleep state or an off state) in order to save their battery power and extend their deployment lifetime, see (Singh & Raghavendra, 1998; Stewart & Tragoudas, 2007; Ye et al., 2004) among others.

PAMAS (Singh & Raghavendra, 1998) is a MAC protocol based on RTS/CTS. PAMAS schedules sleep intervals for sensor nodes to avoid overhearing and uses separate channels for data and control frames. In PAMAS, nodes probe their neighbors for transmission time in order to avoid collision as well. PAMAS reduces energy consumption by avoiding collision and transmission overhearing at the expense of increased hardware complexity, which in turn affects the power consumption.

The S-MAC (Ye et al., 2004) protocol reduces the energy consumption of the nodes by implementing the following mechanisms. First, it reduces idle listening by scheduling sleep intervals for nodes, in fact, S-MAC coordinates sleep intervals amongst neighboring nodes. Second, it divides long messages into smaller packets and transmits them back to back. As a result, nodes with longer messages occupy the wireless medium for longer periods of time. The authors show that this seemingly "unfair" advantage results in energy savings over traditional "fair" methods. Third, it implements a low-duty-cycle that reduces idle listening. Finally, it uses in-channel signaling to reduce overhearing by extending the work from PAMAS (Singh & Raghavendra, 1998).

S-MAC's mechanisms do reduce energy consumption at the expense of increased message latency. However, the predefined sleep intervals limit the flexibility of the protocol and the broadcast mechanism increases the probability for collision because S-MAC does not use RTS/CTS (Demirkol et al., 2006).

TMAC (Van & Langendoen, 2003) is similar to SMAC except that each node is equipped with a timer. In TMAC, a node is put on the low-power/sleep state if it does not transmit or receive for the entire duration of the timeout period. TMAC performs significantly better than S-MAC under variable load.

In WiseMAC (El-Hoiydi & Decotignie, 2004), the authors propose a downlink (to be used when the sink transmits packets to sensors). WiseMAC uses non-persistent CSMA (np-CSMA) with preamble sampling in order to decrease idle listening. In this case, a preamble is used to alert the receiving node that a data packet is on its way. The preamble precedes each data packets. All the nodes in the medium listen to the medium for a constant time interval referred to as the sampling period. If a node hears a transmission while it is listening to the medium, it will continue to listen until it receives a frame or until the medium becomes idle. The sink precedes each data frame with a preamble sequence that is equal to the sampling period. This guarantees that the receiving node will be able to detect the transmission. On the downside,

the long preamble sequence results in a low throughput and in increase power consumption. In addition, all the nodes within wireless range of the receiving node are able to hear the transmission. WiseMAC proposes an improvement to this where the sink takes advantage of knowing the sampling schedule of the nodes. The sink therefore, sends a smaller preamble and a frame right when the receiving node is scheduled to start sampling the medium.

WiseMAC suffers from two main drawbacks (Demirkol et al., 2006). The first drawback results from its decentralized sleep schedule where nodes wake up from their sleep cycle at different times. This is inefficient when broadcast communication is used because the broadcasted frames would need to be stored at the neighbors who are awake and end up being transmitted multiple times. The second drawback of the protocol is the fact that it is vulnerable to the hidden terminal problem where collision can happen at a node if it receives transmissions from two nodes that are not within transmission range of each other (Note that this is not a problem if WiseMAC is only used as a downlink protocol)

TRAMA (Rajendran et al., 2003) is a collision-free TDMA based MAC protocol for sensor networks. TRAMA ensures energy efficiency by avoiding collision during unicast, multicast and broadcast transmissions. In addition, in TRAMA, nodes can switch to a low-power state whenever they are not transmitting or receiving frames to save energy. In TRAMA, a node is elected to transmit within a two-hop neighborhood during each time slot. This mechanism avoids the hidden terminal problem.

TRAMA achieves significant energy savings due to: 1. the increased amount of low-power states, 2. the decreased amount of communication since the receiving nodes are indicated a priori, and 3. the significant decrease in collision probability. However, the latency when using TRAMA is longer compared to CSMA as a result of the high percentage of sleep time (Demirkol et al., 2006).

Berkeley MAC (B-MAC) (Polastre et al., 2004) is a low frills protocol based on clear channel assessment, it uses low power listening with preamble sampling. The default mode in B-MAC does not include a mechanism to avoid the hidden terminal problem, which could be implemented by higher layers if needed. B-MAC achieves significant energy savings when varying check time, by making the preamble constant and setting the sample rate. However, since the protocol is bare-bone, additional features would have to be implemented at higher layers when needed, which increases the complexity of the system as a whole.

Even though multiple MAC layer protocols provide adequate performance, no single protocol has been chosen as a standard. This is due to the fact that some protocols perform better than others for particular applications, communication pattern, network infrastructures and or network densities. An ideal energy efficient MAC layer protocol for WSNs would use a local schedule for motes to turn to the low-power/off state as a function of their residual energy as well as their sensing schedule. The schedule should aim to maximize the sleep time of the motes while preserving their sensing schedule, local connectivity and balancing their energy levels in order to increase the lifetime of the network as a whole.

4.4.1 Physical layer

Frequency detection, generation, modulation and coupling are the responsibility of the physical layer and are explained in detail in the hardware section. Note that when an engineer

is charged with designing a physical layer, propagation effects due to the ambient conditions must be considered.

5. Conclusion and future work

This chapter reviews the hardware architecture of wireless sensor motes, as well as their protocol stack focusing on power considerations at every level. We conclude that because of the diversity in WSN applications, it is very difficult to derive a universal power efficient architecture both in terms of hardware and software.

As far as the hardware components in WSNs, many advances have been made over the last few years. These improvements include more efficient apertures with better directivity and lower VSWR. The sensor element has been made to become more resolute while power management has improved due to the accessibility of more exotic materials for energy storage. The future holds near perfect antenna with nearly a 1:1 VSWR ensuring most of the energy leaving the system goes where it's designed to propagate. Researchers at Purdue University are working toward ensuring optical sensors are near perfectly efficient with negative refractive metamaterials and photon collection efforts.

In terms of the WSN protocol stack, no one protocol has been adopted as a WSN standard, rather each protocol is designed to efficiently serve one or more WSN applications. The power efficiency of protocols has become the number one constraint in almost every layer of the protocol stack. More work is needed to design and develop protocols that are less application specific and still power efficient.

6. References

Kemal Akkaya and Mohamed Younis, *A Survey on Routing Protocols for Wireless Sensor Networks*. Elsevier journal of Ad Hoc Networks, Volume 3, pages 325-349.

I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, *Wireless sensor networks: a survey*. Elsevier Computer Networks, Volume 38, pages 393-422, 2002.

ALERT, www.alertsystems.org

S. Basagni, M. Conti, S. Giordano, Iv. Stojmenovic, *Chapter 11: Energy-efficient Communication in ad hoc Wireless Networks*. Mobile ad hoc networking, Wiley-IEEE Press, 2004.

Bontempi, G., and Le Borgne, Y. (2005). *An adaptive modular approach to the mining of sensor network data*. In Workshop on Data Mining in Sensor Networks, SIAM SDM, Newport Beach, CA, USA, April.

D. Braginsky and D. Estin, *Rumor routing algorithm for sensor networks*. Proceedings of the first workshop on Sensor Networks and Applications (WSNA), Atlanta, GA, October 2002.

Wenshan Cai, and Vladmir Shalaev, *Optical Metamatrials: Fundamentals and Applications*. Springer, 2010.

M. Chu, H. Haussecker, and F. Zhao, *Scalable Information-Driven Sensor Querying and Routing for ad hoc Heterogeneous Sensor Networks*. The International Journal of High Performance Computing Applications, Vol. 16, No. 3, August 2002.

C. Coue, T. Franichard, P. Bessiere, and E. Mazer, *Multi-sensor data fusion using Bayesian programming: An automotive application*. Proceedings of the IEEE/RSJ International Coference on Intelligent Robots and Systems. Vol. 1, Lausanne, Switzerland, pages 141-146.

Crossbow Technologies, *http://www.xbow.com/*

Ilker Demirkol, Cem Ersoy, and Fatih Alagoz, *MAC Protocol for Wireless Sensor Networks: a Survey*. IEEE Communications Magazine, 2006.

A. P. Dempster, *A generalization of Bayesian inference*. J. Royal Stat. Soc., Series B 30, pages 205-247, 1968.

Isabel Dietrich and Falko Dressler, *On the lifetime of wireless sensor networks*. ACM Transactions on Sensor Networks, Vol. 5, No. 1, Article 5, February 2009.

Draft-IETF-MANET-DYMO-mib-04: http://tools.ietf.org/html/draft-ietf-manet-dymo-mib-04

Draft-IETF-MANET-NHDP-15: http://tools.ietf.org/html/draft-ietf-manet-nhdp-15

Draft-IETF-MANET-olsrv2-12: https://datatracker.ietf.org/doc/draft-ietf-manet-olsrv2/

A. El-Hoiydi and J.-D. Decotignie, *WiseMAC: An Ultra Low Power Protocol for the Downlink of Infrastructure Wireless Sensor Networks*. Proceedings of the Ninth IEEE Symposium on Computers and Communication, ISCCÕ04, pages 244-251, Alexandria, Egypt, June 2004.

E. H. Elhafsi, N. MItton, D. Simplot-Ryl, *End-to-End Energy Efficient Geographic Path Discovery With Guaranteed Delivery in Ad Hoc and Sensor Networks*. 19th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Cannes, France: IEEE, 15-18 Septembr 2008, pp. 1-5.

Jeremy Elson, Lewis Girod, and Deborah Estrin, *Fine-grained network time synchronization using reference broadcasts*. Proceedings of the ACM OSDI'02, Boston, MA, December 2002.

Mehdi Esnaashari and M.R. Meybodi. *Data Aggregation in Sensor Networks using Learning Automata*. Wireless Networks 20

S. Ganeriwal, R. Kumar, and M Srivastava, *Timing Sync Protocol for Sensor Networks*. Proceedings of the ACM SenSys'03, 2003.

J.V. Greunen, and J. Rabaey, *Lightweight Time Synchronization for Sensor Networks*. Proceedings of the 2nd ACM International Conference on Wireless Sensor Networks and Applications (WSNA'03), San Diego, CA, September 2003.

S. Grime, H.F. Durrant-Whyte, *Data fusion in decentralized sensor networks*. Control Eng. Practices, 2(5):849-63, 1994.

N. Guilar, A. Chen, T. Kleeburg, and R. Amirtharajah, *Integrated Solar Energy Harvesting and Storage*. Proceedings of the International Symposium of Low Power Electronics and Design (ISLPED'06), October 2006.

I. Gupta, D. Riordan, and S. Sampalli, *Cluster-head election using fuzzy logic for wireless sensor networks*. Proceedings of the 3rd Annual Communcation Networks and Services Research Conference (CNSR'05). IEEE, Halifax Canada, pages 255-260.

Tian He, John Stankovic, Chenyang Lu, and Tarek Abdelzaher, *SPEED: A Real-Time Routing Protocol for Sensor Networks*.

Heinzelman, W., Chandrakasan, A., and Balakrishnan, H. (2000). *Energy-efficient communication protocol for wireless microsensor networks*. In Proceedings of 33rd Hawaii International Conference on System Science (HICSS Ô00), January.

W.B. Heinzelman, A. P. Chandrakasan, and H. Blakrishnan, *An Application-Specific Protocol Architecture for Wireless Microsensor Neworks*. IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 660-670, October 2002.

W. Heinzelman, J. Kulik, and H. Balakrishnan, *Adaptive protocols for information dissemination in wireless sensor networks*. Proceedings of the 5th annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiComÕ99), Seattle, WA, August 1999.

W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, *Energy-Efficient communication protocol for wireless sensor networks*. Proceedings of the Hawaii International Conference System Sciences, Hawaii, January 2000.

Fei Hu, and Xiaojun Cao, *Wireless Sensor Networks: Principles and Practice*. CRC Press, 2010.

C. Intanagonwiwat, R. Govindan and D. Estin, *Directed Diffusion: A scalable and robust communication pradigm for sensor networks*. Proceedings of the 6th annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiComÕ00), Boston, MA, August 2000.

Sandy Irani, Sandeep Shukla, and Rajesh Gupta, *Power Savings* . Proceedings of ACM Transactions on Algorithms, Vol. 3, No. 4, Article 41, November 2007.

Konard Iwanicki, and Maarten van Steen, *On Hierarchical Routing in Wireless Sensor Networks*. Proceedings of IPSN'09, April 13-16, 2009, San Francisco, California, USA.

R. E. Kalman, *A new approach to linear filtering and prediction problems*. Transactions. ASME J. Basic Engin. 82, pages 35-45, 1960.

Mauritus Morne, *Reliable Ad-hoc On-demand Distance Vector Routing Protocol*. Proceedings of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICNICONSMCL'06).

Joanna Kulik, Wendi Rabiner, Hari Balakrishnan, *Adaptive Protocols for Information Dissemination in Wireless Sensor Networks*. Proceedings of the 5th ACM/IEEE Mobicom Conference, Seattle, WA, August 1999.

P. Kyasanur, R. R. Choudhury, and I. Gupta. *Smart Gossip: An Adaptive Gossip-based Broadcasting Service for Sensor Networks*. Proceedings of MASS'06, 2006.

Andreas Lachenmann, Pedro Jos'e Marr'on, Matthias Gauger, Daniel Minder, Olga Saukh, and Kurt Rothermel, *Removing the Memory Limitations of Sensor Networks with Flash-Based Virtual Memory*. Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007.

Xu, Y., Lee, W. C., Xu, J., and Mitchell, G. (2006). *Processing window queries in wireless sensor networks*. In IEEE International Conference on Data Engineering (ICDEÕ06), Atlanta, GA, April.

Christoph Lenzen, Philipp Sommer, and Roger Wattenhofer, *Optimal clock synchronization in network*. Proceedings of the the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys'09).

Christoph Lenzen, Philipp Sommer and Roget Wattenhofer, *Optimal Clock Synchronization in Networks*. Proceedings of SensSysÕ09, November 4-9, 2009, Berkeley, CA, USA.

P. Levis, A. Tavakoli and S. Dawson-Haggerty, *Overview of Existing Routing Protocols for Low Power and Lossy Networks*. IETF ROLL, IETF draft, 14 February 2009, Draft-ietf-roll-protocols-survey-07.

Qun Li, Javed Aslam, and Daniela Rus, *Online power-aware routing in wireless ad-hoc networks*. Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, 2001.

Liqun Li, Guoliang Xing, Limin Sun, Wei Huangfu, Ruogu Zhou, and Hongsong Zhu, *Exploiting FM Radio Data System for Adaptive Clock Calibration in Sensor Networks*. Proceedings of the ACM MobiSys'11, Washington DC, June 28, 2011.

Liu, C., Wu, K., and Pei, J. (2005). *A dynamic clustering and scheduling approach to energy saving in data collection from wireless sensor networks*. In Proceedings of the Second Annual

IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECONÕ05), Santa Clara, California, USA, September.

V. Loscri, G. Morabito, and S. Marano, *A Two-Level Hierarchy for Low-Energy Adaptive Clustering Hierarchy*. Proceedings of the Vehicular Technology Conference (VTC'05), September 25-28, 2005.

Lotfinezhad, M., and Liang, B. (2004). *Effect of partially correlated data on clustering in wireless sensor networks*. In Proceedings of the IEEE International Conference on Sensor and Ad hoc Communications and Networks (SECON), Santa Clara, California, October.

Saoucene Mahfoudh, and Pascale Minet. *Energy-aware Routing in Wireless Ad Hoc and Sensor Networks*, Proceedings of the 6th International Wireless Communications and Mobile Computing Conference (IWCMCÕ10). June 2010.

Alan Mainwaring, Joseph Polastre, Robert Szewczyk, David Culler, and John Anderson. *Wireless Sensor Networks for Habitat Monitoring*. Proceedings of WSNA'02, September 28, 2002, Atlanta, Georgia.

M. Marina, and S. Das, *On-demand Multipath Distance Vector Routing in Ad Hoc Networks*. Proceedings of the 2001 IEEE International Conference on Network Protocols (ICNP), pages 14-23, IEEE Computer Society Press, 2001.

M. Maroti, B. Kusy, G. Simon, and A. Ledeczi, *The Flooding Time Synchronization Protocol*. Proceedings of the 2nd ACN Conference on Embedded Networked Sensor Systems (SenSys'04), Baltimore, Maryland, 2004, pages: 39-49.

Morteza Maleki, Karthik Dantu, and Massoud Pedram *Power-aware Source Routing Protocol for Mobile Ad Hoc Networks*. Proceedings of the ISLPED'02, August 12-14, 2002, Monterey, California, USA.

Eduardo Nakamura, and Alejandro Frery. *Information Fusion for Wireless Sensor Networks: Methods, Models, and Classifications*, Computing Surveys (CSUR), Volume 39, Issue 3, 2007.

Evgenii E. Narimanov, and Alexamder V. Kildishev, *Optical black hole: broadband omnidirectional light absorber*. Appl. Phys. Lett. 95, 041106 (2009).

NTP: http://www.ntp.org/

OLSR: RFC 3626

D. Petriovic, R. C. Shah, L. Ramchandran, and J. Rabaey, *Data funneling: Routing with aggregation and compression for wireless sensor networks*. Proceedings of the first IEEE International Workshop on Sensor Network Protocols and Applications (SNPA'03). IEEE, Anchrage, AK, pages 156-162.

Kris Pister, *Autonomous sensing and communication in a cubic millimeter*. Http://www-bsac.eecs.berkeley.edu/ pister/SmartDust/

Joseph Polastre, Jason Hill, and David Culler, *Versatile low power media access for wireless sensor networks*. Proceedings of SenSys'04, 2004.

Octavian Postolache, Pedro Silva Girao, and Jose Miguel Dias Pereira, *Non-Volatile Memory Interface Protocols for Smart Sensor Networks and Mobile Devices*. Data Storage, InTech publishers, April 2010.

V. Rajendran, K. Obraczka, J. J. Garcia-Luna-Aceves, *Energy-Efficient, Collision-Free Medium Access Control for Wireless Sensor Networks*. Proceedings of SenSysÕ03, November 5-7, 2003, Los Angeles, California, USA.

RFC 1142: http://tools.ietf.org/html/rfc1142

RFC 2328:http://www.ietf.org/rfc/rfc2328.txt

RFC 2453: http://tools.ietf.org/html/rfc2453

RFC 3561: http://www.ietf.org/rfc/rfc3561.txt
RFC 3626:http://www.ietf.org/rfc/rfc3626.txt
RFC 3684: http://www.ietf.org/rfc/rfc3684.txt
RFC 4728: http://www.ietf.org/rfc/rfc4728.txt
RFC 4861: http://tools.ietf.org/html/rfc4861

Kay Romer, *Time synchronization in ad hoc networks*. Proceedings of the ACM MobiHoc'01, Long Beach, CA October 2001.

Rosemark, R., and Lee, W. C. (2005). *Decentralizing query processing in sensor networks*. In The Second International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiquitousÕ05), San Diego, CA, July (pp. 270Ð280).

C. Schurgers and M.B. Srivastava, *Energy efficient routing in wireless sensor networks*. MILCOM Proceedings on Communications for Network-Centric Operations: Creating the Information Force, McLean, VA 2001.

G. Shafer, *A Mathematical Theory of Evidence*. Princelton University Press, Princeton, NJ 1976.

Vladimir M. Shalaev, Wenshan Cai, Uday K. Chettiar, Hsiao-Kuan Yuan, Andrey K. Sarychev, Vladimir P. Drachev, and Alexander V. Kildishev, *Negative index of refraction in optical metamaterials*. Optics Letters, Vol. 30, Issue 24, pages 3356-3358, 2005.

S. Singh and C.S. Raghavendra, *Power aware multi-access protocol with signaling for ad hoc networks*. ACM Computer Communication Review Vol. 28 No. 3 (July 1998) pages. 5-26

Philipp Sommer and, Roger Wattenhofer, *Gradient Clock Synchronization in Wireless Sensor Networks*. Proceedings of the International Conference on Information Processing in Sensor Networks, 2009.

Soro, S., and Heinzelman, W. (2005). *Prolonging the lifetime of wireless sensor networks via unequal clustering*. In Proceedings of the 5th International Workshop on Algorithms for Wireless, Mobile, Ad Hoc and Sensor Networks (IEEE WMAN Ô05), April.

K. Stewart, and S. Tragoudas, *Managing the power resources of sensor networks with performance considerations*. Computer Communications Journal, Volume 30, Number 5, pages:1122-1135, March 2007.

A. Tsymbal, S. Puuronen, and D. W. Patterson, *Ensemble feature selection with the simple Bayesian classification*. Information Fusion 4, 2, June, pages 87-100.

T. van Dam, and K. Langendoen, *An adaptive energy-efficient mac protocol for wireless sensor networks*. Proceedings of the First ACM Conference on Embedded Networked Sensor Systems, November 2003.

Natalia Vassileva, Francisco Barcelo-Arroyo, *A Survey of Routing Protocols for Maximizing the Lifetime of Ad Hoc Wireless Networks*. International Journal of Software Engineering and its Applications, Vol. 2, No. 3, July 2008.

Virrankoski, R., and Savvides, A. (2005). *TASC: Topology adaptive spatial clustering for sensor networks*. In Second IEEE International Conference on Mobile Ad Hoc and Sensor systems, Washington, DC, November.

Thomas Watteyne, Maria Grazia Richichi, *From MANET to IETF ROLL Standardization: A Paradigm Shift in WSN Routing Protocols*, submitted to IEEE Communications Surveys and Tutorials.

Jon Wilson, *Sensor Technology Handbook*. Elsevier, ISBN: 0-7506-7729-5, December 2004.

Winter, J., Xu, Y., and Lee, W. C. (2005). *Energy efficient processing of K nearest neighbor queries in location-aware sensor networks*. In The Second International Conference on Mobile and

Ubiquitous Systems: Networking and Services (MobiquitousÕ05), San Diego, CA, July (pp. 281Ð292).

L. Xiao, S. Boyd, and S. Lall, *A scheme for robust distributed sensor fusion based on average consensus*. Proceedings of the 4th International Symposium on Information Processin in Sensor Networks (IPSN'05), pages 63-70, 2005.

Z. Xiong, A. D Liveris, and S. Cheng, *Distributed source coding for sensor networks*. Proceedings of IEEE Sig. Proc. Mag. 21, 5, September 2004, pages 80-94.

M. Ye, C. Li, G. Chen, and J. Wu, EECS: *An Energy Efficient Clustering Scheme in Wireless Sensor Networks*. Ad Hoc and Sensor Wireless Networks, Vol. 3, Pages 99-119, April 2006.

W. Ye, J. Heidemann, D. Estrin, *Medium Access Control With Coordinated Adaptive Sleeping for Wireless Sensor Networks*. IEEE/ACM Transactions on Networking, Volume 12, Issue: 3, Pages: 493-506, June 2004.

Zhenzhen Ye, Alhussein Abouzed, and Jing Ai. *Optimal Stochastic Policies for Distributed Data Aggregation in Wireless Sensor Networks*, IEEE/ACM Transactions on Networking, VOL. 17, NO. 5, October 2009.

O. Younis and S. Fahmy, *HEED: A Hybrid Energy-Efficient Distributed Clustering Approach for Ad Hoc Sensor Networks* IEEE Transactions on Mobile Computing, vol. 3, no. 4, Oct-Dec 2004.

Younis, O., and Fahmy, S. (2005). *An experimental study of routing and data aggregation in sensor networks*. In Proceedings of theInternational Workshop on Localized Communication and Topology Protocols for Ad hoc Networks (LOCAN), held in conjunction with The 2nd IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS-2005), November.

7

Review of Optimization Problems in Wireless Sensor Networks

Ada Gogu[1], Dritan Nace[1], Arta Dilo[2] and Nirvana Meratnia[2]
[1]*Université de Technologie de Compiègne*
[2]*University of Twente*
[1]*France*
[2]*The Netherlands*

1. Introduction

Wireless Sensor Networks (WSNs) are an interesting field of research because of their numerous applications and the possibility of integrating them into more complex network systems. The difficulties encountered in WSN design usually relate either to their stringent constraints, which include energy, bandwidth, memory and computational capabilities, or to the requirements of the particular application. As WSN design problems become more and more challenging, advances in the areas of Operations Research (OR) and Optimization are becoming increasingly useful in addressing them.

This study is concerned with topics relating to network design (including coverage, topology and power control, the medium access mechanism and the duty cycle) and to routing in WSN. The optimization problems encountered in these areas are affected simultaneously by different parameters pertaining to the physical, Medium Access Control (MAC), routing and application layers of the protocol stack. The goal of this study is to identify a number of different network problems, and for each of these network problems to examine the underlying optimization problem. In each case we begin by presenting the basic version of the network problem and extend it by introducing new constraints. These constraints result mainly from technological advances and from additional requirements present in WSN applications. For all the network problems discussed here a wide range of algorithms and protocols are to be found in the literature. We cite only some of these, since we are concerned more with the network optimization problem itself, together with its different versions, than with a state of art of methods for solving it. Moreover, the cited methods have originated in a variety of disciplines, with approaches ranging from the deterministic to the opportunistic, including computational geometry, linear, nonlinear and dynamic programming, metaheuristics and heuristics, game theory, and so on. We go on to discuss the complexity inherent in different optimization problems, in order to give some hints to WSN designers facing new but similar scenarios. We try to highlight distributed solutions and information that is required to implement these schemes. For each topic the general presentation scheme is as follows:

i) Present the network problem

ii) Identify the relevant optimization problem

iii) Discuss the theoretical complexity of the optimization problem

iv) Describe some representative solution methods, including distributed methods

The relations between the two areas of WSN network design and OR have been discussed in some other works (Li, 2008; Nieberg, 2006; Ren et al., 2006; Suomela, 2009). In (Li, 2008; Nieberg, 2006; Suomela, 2009) the goal is to relate a network problem to its corresponding optimization problems and to discuss related questions in the OR literature that might feature in a solution. For example, Suomela (2009) is focused on data gathering and scheduling problems in WSN. He identifies the respective optimization problems and presents some nice properties that a graph should have (e.g. bipartite, graph with unique identifiers, planar, spanners, etc) to facilitate the design of distributed algorithms for these optimizations problems. Ren et al. (2006) present a survey highlighting certain methodologies from operational research and the corresponding network problems that they can solve. In particular they relate *graph theory and network flow problems* to routing problems in WSN, *fuzzy logic* to clustering, and *game theory* to the problem of bandwidth allocation. Following on from these works we attempt to enlarge the spectrum of the network problems addressed, and for each network problem we highlight the optimization problem together with some effective methods proposed in the literature. Furthermore, we report at the end the study a discussion on open issues.

This chapter is organized as follows. The second section introduces certain methods from OR which are used to solve problems in WSN. The goal is to familiarize the reader with both the terminology and methods that are encountered in the OR domain and we refer to in the reminder of the study. In the third section we discuss several problems of WSN design, most of which must be addressed in the setup phase of the network. The fourth section is concerned with the routing problems. We report a classification of most used models and focus on how each of them is useful in addressing routing problems in WSN. The final section identifies some open issues in WSN and gives concluding remarks.

2. Operations research methodology used in WSN design

This section aims to introduce the reader to OR terminology and some representative solution methods from OR that are already used in WSN design. An Optimization Problem (OP) in OR is composed of two main parts. One is the objective/cost function to be maximized/minimized, and the second is concerned with the associated constraints that determine the feasibility domain. A solution of the OP is feasible if it satisfies all the constraints. From computational complexity point of view an OP is said to be polynomial if there exists a polynomial-time algorithm for solving it, otherwise it falls into NP-hard problems class. The solution methods used to solve the OP can be classed into two groups: exact methods and heuristic methods.

1. **Exact methods** seek a global optimal solution (if it exists) for the problem. The most familiar techniques among the exact methods commonly used for OPs in WSN are Linear, Nonlinear and Dynamic Programming. A general linear programming (LP) formulation is as follows:

$$\max\ cx \tag{1}$$

$$Ax \leq b \tag{2}$$

$$x \geq 0 \tag{3}$$

where A is a matrix, b and c are vectors giving respectively the right-hand terms and the cost coefficients, and x is the decision variable vector.

In cases where some decision variables have integer values while others have continuous values we refer to the problem as *Mixed Integer Linear Programming*. If, on the other hand, the vector x contains only integer values, then we have a case of *Integer Linear Programming (ILP)*. Note that the difficulty of the problem increases when we are dealing with ILP rather than LP, since ILP problems are commonly NP-hard. The most frequently used algorithms for solving LP problems are Simplex and Interior Points methods (Dantzig, 1963; Karmarkar, 1984), whereas for ILP problems there are Branch-and-Bound, Branch-and-Cut and Cutting Planes methods. Besides maximizing/minimizing an objective function, LP can be adapted so that it also guarantees fairness. In this case the objective function becomes a *max-min (or min-max)* objective function. In WSN we may often encounter network problems modeled according to this structure. It also happens that in some networks modeled by LP the number of variables is infinite or finite but huge, making an explicit enumeration impossible. In these cases the problem is solved iteratively. At each iteration new variables that potentially would lead to better solutions are generated by a method called column generation. The problem is solved when no new variables can be generated. Finally, when the objective function, or at least one of the constraints, is a nonlinear function, the problem becomes a nonlinear programming problem. In this type of problem the nature of the objective function is very important. If it is a convex function, then the problem is a nonlinear convex programming problem, where the best-known techniques include subgradient and Lagrangian decomposition (Kuhn, 1951; Shor, 1985). The above linear programming problems can be solved using a commercial solver such as CPLEX, Xpress-MP, etc. For nonlinear nonconvex programming the optimization becomes difficult and the solution methods less tractable. Another method worth citing is *Dynamic programming* (Bellman, 1957). This is a sequential approach where the decisions are taken optimally, step-by-step, until the complete solution has been constructed. This method works for problems that can be divided into subproblems that are simpler to solve and whose solutions will produce the global solution.

2. **Heuristic methods** are an important class of solution methods for practical optimization problems in WSN exhibiting high computational complexity. These approaches are intended to quickly provide near-optimal solutions to difficult optimization problems that cannot be solved exactly. Their advantages include easy implementation, rapidly-obtained solutions and robustness to variations in problem characteristics. However, in most cases they cannot guarantee the quality of the solution produced. Heuristic methods include local improvement methods that perform searches within the neighborhood of a feasible solution to the problem, and improve/construct the solution step by step by taking the best local optimal decision at each step. The main danger here is getting trapped at a local optimum, and to overcome this danger these methods may be combined with random approaches, multi-start approaches, and so on.

   Similarly, metaheuristics are very general approaches used to guide other methods or procedures towards achieving reasonable solutions. Metaheuristics aim at reducing the search space and avoiding local optima. Most metaheuristics are life-inspired approaches such as Tabu Search (Glover, 1989), Evolutionary/Genetic algorithms (EA/GA) (Holland, 1975), Memetic algorithms (Moscato, 1999), Ant Colony Optimization (ACO) (Dorigo et al., 1996), and Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1995). Tabu Search starts with one feasible solution and constructs its neighborhood out of members that

are obtained by permuting the elements of the feasible solution. The objective function is next calculated for each member of the neighborhood and the best one is selected. The process is then repeated but with the newly selected member as the starting point. An important element in this algorithm is loop-avoidance, meaning that it must not return to a solution that has been already processed, and for this reason all the forbidden movements are saved in a tabu list. In evolutionary or genetic algorithms the solutions of the problem are called individuals. A relatively small set of individuals selected within the enormous search space of the optimization problem are chosen to form the population. The population evolves during the iterations in a certain order known as generations. Genetic operators such as mutation and crossover are applied to produce better individuals. Their performance is evaluated based on a fitness/cost function. The algorithm stops when the solution is close to the optimum, or when a specific number of generations has been reached. Memetic algorithms combine GA with a local search. These algorithm follow the logic of a GA, but before applying genetic operators, every individual carries out a local search with the aim of improving its fitness. In ant colony optimization, an ant starts from a random node in the graph and selects the next node based on Equation (4).

$$P_{ij} = \begin{cases} \frac{\tau_{ij}^{\alpha} \cdot \eta_{ij}^{\beta}}{\sum_{k \in Liste} \tau_{ik}^{\alpha} \cdot \eta_{ik}^{\beta}} & \text{if } j \in List, \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where P_{ij} is the probability of choosing node j when the current node is i, τ_{ij} is the pheromone value of edge (i, j), η_{ij} the heuristic value, List contains all possible nodes accessible by the ant, and α, β are constants whose values depend in some way on the nature of the problem. In order to use an ACO algorithm for an OP, it is really important to present meaningfully the pheromone and heuristic values. When an ant passes through a node/edge, it deposits a pheromone value τ_{ij} in it. This value has to be proportional to the quality of the solution and it will help to attract other ants from the colony. The intention is that all the ants end up following the same trail, which hopefully represents the optimal solution. In order to avoid local optima this algorithm contains a process known as evaporation which periodically reduces the pheromone value deposited on a trail. The PSO algorithm imitates the flocking of birds. It initializes a number of agents (birds) and attaches two parameters, position and velocity (the velocity is given by two vectors which have orthogonal directions), to each of them. At each iteration the algorithm has to evaluate the positions of the agents and determine the subsequent positions, while accelerating their movement toward "better" solutions.

3. Network design issues

WSN design has to address a number of challenging factors. These include node deployment and coverage, connectivity and fault tolerance. The overall aim is always to lower costs, reduce the power consumption of the wireless environment and ensure a reliable network. Node deployment is the first essential stage in WSN design, and it strongly impacts the performance of the network as regards accurate event detection and efficient communication. Once the node is deployed, the problems of network organization become crucial, with topology and power control problems on one side, and medium access and scheduling strategies on the other. Solving these problems is an integral part of the design of a viable, energy-efficient network.

3.1 Optimal sensor deployment and coverage

WSN applications have particular requirements to satisfy, and one in common for all of them is coverage. The problem of maximizing the coverage of a given monitoring area has received a lot of attention in the literature. In this subsection we focus on three main problems related to this topic. First we discuss the problem of the minimum number of sensors required to cover a given area and guarantee network connectivity. The second problem is finding the best locations for a finite number of sensor nodes when seeking to satisfy the requirement of event detection. The third problem is identifying the regions that are not covered by sensors, assuming that the deployment is known.

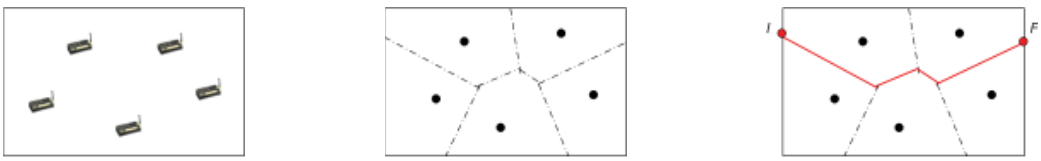
The WSN deployment (or layout) problem is concerned with minimizing the number of deployed sensor nodes while ensuring the full coverage and connectivity of the monitoring area. As presented in (Efrat et al., 2005), the problem is a version of the Art Gallery problem, which is known to be NP-hard. The Art Gallery problem involves placing the smallest number of guards in an area such that every point in it can be surveyed by the guards. In this work Efrat et al. (2005) also show that the problem of deciding whether k sensors are sufficient to survey a region such that every point within the region is covered by three sensors is NP-hard. They propose an approximation algorithm based on geometry calculations for solving the problem.

However, most of the algorithms proposed for the layout problem derive from metaheuristic and heuristic methods. The work of Rotar et al. (2009) uses a new algorithm known as the Guided Hyper-plane Evolutionary Algorithm (GHEA). GHEA behaves basically the same as a multi-objective evolutionary algorithm manipulating a population and individuals. Whereas in the evolutionary algorithms the individuals are evaluated according to a fitness function, the novelty of GHEA lies in its evaluation of the population based on the hyperplane. The hyperplane will consist of points in the space which have better performances than any individual within the current population. Fidanova et al. (2010) propose an ant colony algorithm for this problem. As previously mentioned, ACO algorithms emulate the behavior of real ant colony where the greater the number of ants following a trail, the more attractive the trail becomes. In this case the area is modeled as a grid and all the points on the grid (or nodes) represent the search space. In order to apply the ant algorithm for the layout problem, from Equation (4) it is necessary to calculate the pheromone and the heuristic value every time that an ant passes through a node. The heuristic value attempts to reflect the best candidate node for the future movement of the ant (the new sensor placement) based on local information such as the number of grid points that the new candidate covers, whether the new candidate is reachable at a given distance, which is determined by the sensor transmission range, and finally whether this new position has already been selected by another ant. The pheromone, on the other hand, is initialized with a small value (e.g. the inverse of the number of ants) and for the upcoming iterations its value is updated according to the best solution value of the previous iteration.

In terms of the quality of service, attempts are made to find areas of lower observability from sensor nodes and to detect breach regions. The problem known as the Sensor Location Problem (SLP), formulated by (Cavalier et al., 2007), can be stated as follows: given a planar region, a given number of sensor nodes need to be positioned so that the probability of detecting an event in this region is maximized. The non-detection probability is expressed as a function of the distance between the sensor and a given position in the space where an

event may take place, while the objective function aims to minimize the maximum of this product. In this formulation the problem is a difficult nonlinear nonconvex programming problem. Cavalier et al. (2007) proposes a heuristic algorithm that uses Voronoi polygons to estimate the probability of non-detection and to determine a search direction. The heuristic begins with an initial solution of m sensor locations $(x_1, x_2, ...x_m)$, on the basis of which the Voronoi diagram is constructed (see Fig. 1(a) and (b)). The construction of the Voronoi diagram must also take into account the area of the region. For every node the algorithm determines the point in the Voronoi polygon with the highest probability that an event will not be detected, and defines these points as the new node locations. The process is repeated until no further improvement is possible. We note that a similar problem is encountered by the wireless communication community in GSM networks and content-distribution wired networks (CDN). In GSM networks the problem is to find an optimal deployment of base stations within a region so that it provides maximum possible coverage. In CDN the problem is to determine the locations of proxies where the popular streams can be cached. This problem turns out to be the classical weighted $p - center$ location problem, where the objective is to locate p identical facilities that minimize the maximum weighted distance between clients and their corresponding (closest) facilities, assuming that each client is served by the closest facility (Averbakh & Berman, 1997). The $p - center$ problem is slightly different from SLP (note that for the SLP problem the clients correspond to events and the facilities correspond to sensors). A $p - center$ solution gives an assignment because each demand is assigned to a facility, while in SLP the event point (demand) can be visible to more than one sensor node (facility).

Once the sensors are deployed, coverage describes how well the sensors observe their target area or certain moving targets within this target area. In this context we need to know the path, known as the maximal breach path, that minimizes the maximum distance between every point on the path and its nearest sensor node. In other words this path represents the shortest path connecting the two endpoints which remains as far away as possible from sensor nodes. It was shown in (Duttagupta et al., 2008) that this problem is NP-hard. Most works in the literature propose methods relying on computational geometry and graph theory. Meguerdichian et al. (2001) suggest constructing the Voronoi diagram for the set of nodes in order to compute the maximal breach path. The edges of the Voronoi diagram provide the points of space which are at the greatest distance from the given set of sensors. These edges are weighted according to their distance from the nearest sensor. In this graph, the maximal breach path is a path maximizing the weight of its edges. A breadth-first-search (BFS) algorithm is then applied to find the maximal breach path. The Voronoi diagram and the maximal breach path are depicted in Fig. 1.



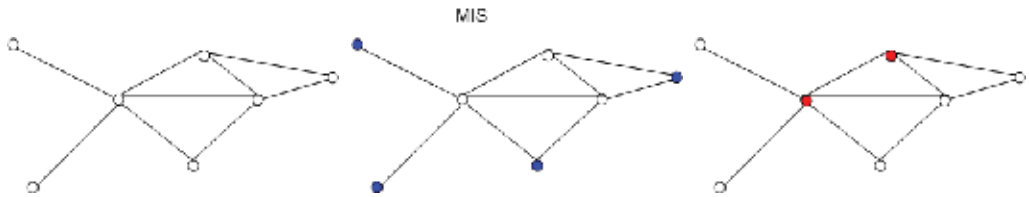
(a) A WSN (b) Voronoi diagram (c) Maximal breach path IF

Fig. 1. Voronoi diagram (b) for WSN nodes shown in (a), and maximal breach path (c).

3.2 Topology control

Node deployment can give rise to dense networks where sensors can have multiple potential neighboring nodes in common. This situation may lead to congestion and energy waste. To overcome this problem, topology control techniques are used to reduce the initial topology by choosing a subset of nodes having some property. Here the problem is finding a strongly connected subset of nodes that covers the rest of the nodes, so as to guarantee the connectivity of the whole network. This subset will be the backbone of the network, and every node excluded from it must have at least one edge in common with a node belonging to the subset. There are a number of advantages in obtaining a backbone topology, since for instance it may i) reduce network traffic by performing data aggregation and in-network processing, ii) avoid packet collisions as only the backbone nodes will forward packets to the sink while improving network throughput, and iii) make it possible to turn off the non-backbone nodes to save energy. This subsection will discuss the optimization problem for constructing the reduced topology, and the special case in which the lossy links are taken into account.

The problem is modeled as a widely-known mathematical problem called the Connected Dominating Set (CDS). A Dominating Set of a graph $G(V_{nodes}, E_{edges})$ is the subset of nodes $D \subset V$, such that every node that does not belong to D has at least one link in common with a node in D. In the special case in which these nodes have to be connected, the set is called the Connected Dominating Set (CDS). For many applications the smallest dominating set is sought, which brings us to the problem of finding the Minimum Connected Dominating Set (MCDS). The nodes in a CDS are called dominators, while other nodes are called dominatees. The MCDS problem is known to be NP-hard and is of the same difficulty and directly convertible to the vertex cover problem, the independent set computation problem, or the maximum clique problem. Yuanyuan et al. (2006) propose a two-phase method for obtaining



(a) A graph (b) MIS (c) MCDS

Fig. 2. Construction of the Minimum Independent Set - MIS (b) and Minimum Connected Dominating Set - MCDS (c) of the graph shown in (a)

the CDS. In the first phase a Maximal Independent Set (MIS) is formed. An Independent Set (IS) of a graph G is the node subset S where no two nodes in S have an edge in common. The MIS is the maximal IS, which means that it is not possible to include more nodes in S. In the second phase, the goal is to build a CDS using nodes that do not belong to the MIS. These nodes are selected in a greedy manner. At the end, the non-MIS node with the highest weight (the weight depends on the remaining energy and the degree of the node) becomes part of the CDS, as depicted in Fig. 2. Unfortunately, a CDS only preserves 1-connectivity and it is therefore very vulnerable. When fault tolerance against node failures is taken into account, the problem becomes the $kmCDS$ problem, (k-Connected m-Dominating Set). The requirement of $k - connectivity$ guarantees that between any pair of dominators there exist at least k different paths, and the $m - domination$ guarantees that each dominatee is connected

with m dominators. Wu & Li (2008) propose a distributed algorithm for this problem with time complexity $O((m + \Delta) \cdot Diam)$, where Δ is the maximum node degree and $Diam$ is the diameter of the network (the length of the longest shortest path between any pair of nodes in the graph). Li (2008) assumes that the MCDS nodes are aligned according to a strip-based deployment pattern, as in Fig. 3 where the nodes are deployed in straight lines. The difference with a grid pattern is that the odd lines are horizontally shifted by a given distance in relation to the even lines. This pattern is shown to be a near-optimal solution of MCDS for an infinite network in terms of space. Because a WSN is a finite network, the spacing parameter in this pattern and consequently the number of nodes needs to be adapted. The optimization problem aims to minimize the number of nodes in the strip-based pattern such that the areas, defined by the node's transmission range, of three neighboring nodes in this pattern intersect each other (see Fig. 3). The solution of this problem gives the positions of the CDS. Implementation in a real scenario is easier. Assuming a given finite area with the sensor nodes uniformly deployed in it, for every position determined by the algorithm the closest sensor in the network will be selected for belonging to the CDS. The distributed approach requires

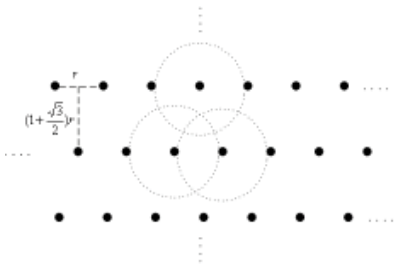
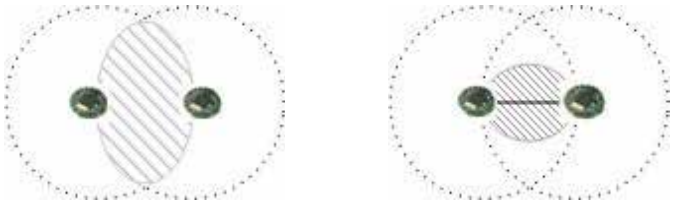


Fig. 3. A strip-based pattern (redrawn from (Li, 2008)).

that nodes exchange certain information, such as the distance from the ideal positions and the number of neighbors that they cover, in order to make a decision regarding membership of the CDS. Nonetheless, the problem of finding the MCDS becomes more complex for dynamic or mobile networks, and this question is still open.

Up to now we have taken "neighbors" to refer to those nodes that are reachable if a node transmits with a given power. In (Liu et al., 2010; Ma et al., 2008) the authors also take into account the existence of *lossy links*. A lossy link has an additional parameter representing the probability of a successful transmission over the link. Topology control algorithms that consider these links are known as *opportunistic* algorithms. The related problem in (Ma et al., 2008) is to minimize the number of hops between a node in the network and the sink while guaranteeing that the path utility (the utility used is a metric reflecting the expected number of packet transmission required to successfully deliver a packet) falls within in a given interval. The distributed approach requires that a node knows the utility value and the IDs of its $2 - hop$ neighbors and that it decides whether or not to act as a relay node. (Liu et al., 2010), on the other hand, demonstrated that the problem of finding a subnetwork of the original network (the subnetwork has to contain all the nodes but only a subset of link of the original network) which minimizes the overall energy consumption and guarantees that the *reachability coefficient* (RC) for every node-sink pair exceeds a particular threshold is NP-hard. RC is a coefficient that indicates the probability of a node being able to reach another node in the network while the respective threshold is imposed by application requirement. When calculating the RC for two nodes that are connected by a path, the RC will be equal to the



(a) Intersection zone for RNG

(b) Intersection zone for GG

Fig. 4. Intersection zones for constructing Relative Neighbor (RNG) and Gabriel (GG) graphs.

mean of the RC values of the links that constitute the path. The key idea of their solution is that link-disjoint trees can be constructed, the union of which will give the subnetwork. A node will make a decision to join in some tree construction if its RC is less than a particular threshold.

3.3 Power control

Unlike the topology control problem which seeks to minimize the size of the network backbone while assuming uniform and constant power transmission, the power control problem (also referred to as the Range Assignment (RA) problem or Strong Minimum Energy Topology (SMET)) aims to fix the node's transmission power at appropriate levels. The goal here is to reduce energy consumption while preserving connectivity in the network. Different methods proposed in the literature for solving this problem are discussed in this subsection. We present some extended versions which add new constraints to this problem with respect to i) throughput, ii) traffic and iii) reliability.

SMET has been shown by Cheng et al. (2003) to be an NP-hard optimization problem. To tackle the problem they propose two heuristics: Minimum Spanning Tree (MST), where power is assigned to nodes such that they can reach the farthest children in the MST, and Incremental Power (IP). In the IP heuristic the power of the node is allocated in a greedy manner. The heuristic begins with an empty set of nodes, to which it then adds a node chosen randomly from the network. This node adjusts its power to reach its closest neighbor. Further, each member of the set tries to increase its power to include another node, but the only member to succeed will be the one that expends the least energy in achieving this end. The algorithm stops when all the nodes are included in the set. Since transmitting with the same power can lead to energy waste, some methods based on computational geometry, such as Relative Neighbor Graph (RNG) (Wan et al., 2001), Gabriel Graph (GG) (Ke et al., 2009), Yao graph or Voronoi Diagram have been put forward to determine the "best neighborhood". In these methods two nodes can be neighbors if there are no other nodes in the zone of intersection. The main difference between them is the way that they define this intersection zone. Fig. 4 shows how the intersection zone is constructed in RNG and GG. The idea behind computational geometry implementations is that the energy cost of transmitting directly to some nodes would be less than the cost of using any other relaying scheme to reach them, and so it is worthwhile to use certain methods to discover a node's best neighbors. In many cases the node can reduce its energy so as to be able to reach only its best neighbors. Many algorithms proposed to construct these graphs are centralized, but there also exist distributed versions (Li et al., 2002). A memetic algorithm is proposed by Konstantinidis et al. (2007)

to solve the SMET problem. In reality, the difficulty of applying this kind of algorithm is modeling the problem according to the algorithm's logic, and deciding for example how to define a chromosome, how to implement crossover, how to handle population diversity, etc. The solution to the SMET problem takes the form of an array of positive integers, in which the elements of the array correspond to the power levels assigned to each node, and the respective indexes correspond to the node ID. In Fig. 5 we have 5 sensor nodes which are transmitting with a given power. From this scenario an array of 5 elements is constructed which contains the power values ordered by node ID. This array alternatively represents an

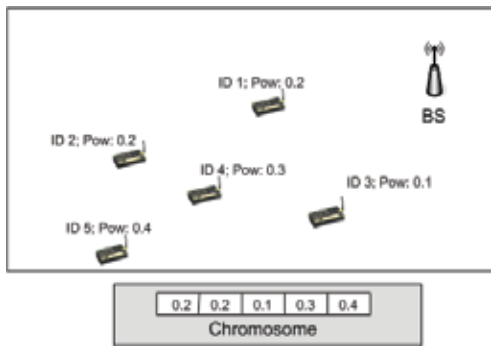


Fig. 5. Chromosome Construction

individual, a chromosome, or a solution, depending on the point of view. The objective of the SMET problem is given by the fitness function defined by the sum of the powers assigned to each node. The first phase of the algorithm proceeds by initializing a random population. It then applies a local search to check the feasibility of the solutions, modifies them in order to obtain feasible ones, and improves the solutions by reducing the assigned power if it is possible. In the second phase a genetic algorithm is applied which involves the crossover of the selected individuals and the mutation for maintaining population diversity. Finally the best individuals from each generation are generated. The procedure is repeated until the solution cannot be further improved.

Lately, this problem has been extended to take some other important parameters into account. The problem of maximizing *the throughput* using topology control is discussed in (Tao et al., 2010). Assuming that the WSN is presented through an RNG (or a GG), their algorithm adjusts the intersection zone between two neighbor nodes in the respective graph (the intersection zone between two neighbors is depicted in Fig. 4) such that the throughput is maximized. They show that if the area of the intersection zone between two neighboring nodes changes in a given interval, the network will preserve the connectivity and energy efficiency properties. Their solution is based on mathematical analysis and a complex equation is derived to find the optimal solution which guarantees the maximal throughput. The equation takes as inputs the node density and the expected throughput of the network. In Gogu et al. (2010), on the other hand, there is a discussion concerning the problem of transmission range assignment and optimal deployment to reduce the energy consumption while taking node traffic into account. The solution is based on dynamic programming methods and it gives the optimal number of sensor nodes and their transmission ranges for a linear network operating under different traffic scenarios. This work also includes an extension to the multihop network case with aggregation (Fig. 6). Hence for a given random deployment of sensors (the blue points in figure), the algorithm calculates the number of nodes that will be in charge for aggregating and

relaying the data towards the base station (the red points), their location, and the respective transmission range. Valli & Dananjayan (2008) discuss the problem of topology control to

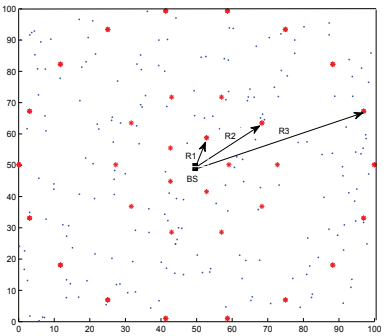


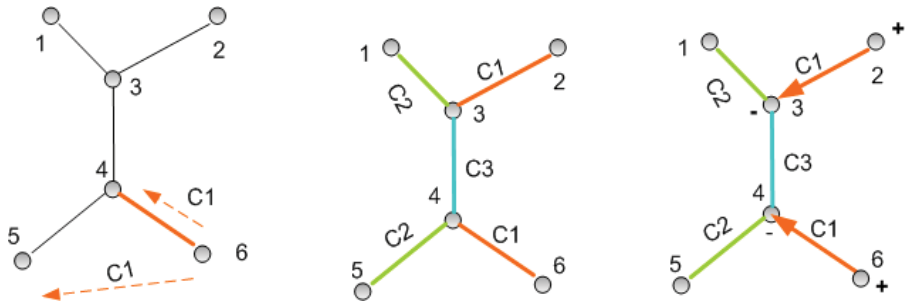
Fig. 6. Optimal position of sensors (red points) in a random deployment.

maximize *network reliability* measured by the bit error rate (BER). They model the problem as a game where a node in the network represents a player. Based on some local information a node calculates the utility function which depends on the link's BER. In every iteration each node will try to optimize this function in a non-cooperative way until the system reaches Nash equilibrium. Another approach is adopted by (Yang & Cai, 2008) to deal with *QoS requirements*. Residual energy, end-to-end delay and link loss ratio are the QoS parameters considered. The question is how to allocate the power values to the nodes such that the energy consumption is minimized, the network is connected and the above QoS requirements are met. The solution proposed is a distributed heuristic based on the minimum spanning tree (MST), where the link metric is a function of delay and packet loss ratio. When this tree is constructed, a node adjusts its power simply so as to be able to reach its parent.

3.4 Medium access strategies

In this subsection we are looking at a node's strategies for accessing the medium. These strategies govern the coordination between the nodes in the network in order for them to access the medium and perform successful transmissions. Most of the work related to medium access strategies in WSN is related to the two main approaches, which are, first, scheduled and secondly, random access/contention-based (Ye & Heidemann, 2003). TDMA (Time Division Multiple Access) is one of the common mechanisms falling under the scheduled approaches, whereas CSMA (Carrier Sense Multiple Access) and derivatives are the most commonly-used methods based on channel contention. Other solutions, more common in cellular networks, but also used by the WSN community, are FDMA (Frequency Division Multiple Access) and CDMA (Code Division Multiple Access). The TDMA, FDMA and CDMA mechanisms are employed in WSN to ensure a collision-free medium access. In this subsection we present the basic problem related to each of them. We then describe three extended versions of TDMA relating to i) connectivity, ii) traffic and iii) delay. For FDMA the extended constraint is throughput. Regarding CDMA, we discuss the problems related to joint use of CDMA with TDMA or FDMA.

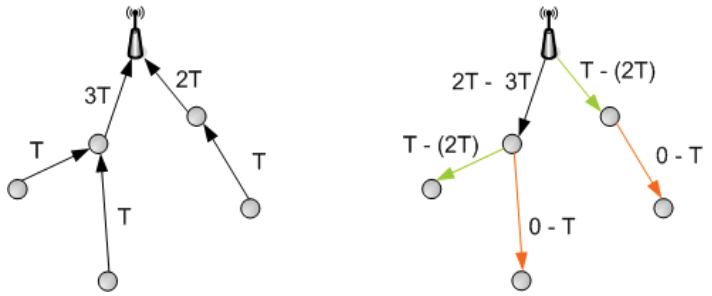
Under the scheduled approach the basic problem is to obtain a slot allocation for all nodes in the network using the smallest possible number of slots such that k-hop neighbor nodes (where k is a positive integer usually equal to 2) are not allocated to the same time slot. The respective optimization problem is the chromatic graph optimization problem, which aims to minimize the number of colors used to color the nodes such that two neighbor elements do not use the same color. This problem is addressed in several works that have put forward a number of distributed algorithms (Al-Khdour & Baroudi, 2010; Gandham et al., 2005; Kawano & Miyazaki, 2009; Sridharan & Krishnamachari, 2004). In (Sridharan & Krishnamachari, 2004) slot allocation uses the logic of a breadth-first search algorithm where the first node which allocates the slot is the root of the tree (the sink). Once a node is selected it continues the operation of slot allocation based on the information from its neighbors. Gandham et al. (2005) discuss the edge-coloring problem, where two edges incident on the same node cannot be assigned to the same time slot. They propose a greedy heuristic whose first step involves coloring the edges and whose second step proposes a strategy to map the colors to the time slots. The second step uses the edge orientation to avoid the hidden and exposed node terminal problem. A simple example is shown in Fig. 7. The process begins with node 6 (the node with the largest ID), which picks a color from a set of colors and broadcasts this information to its neighbors. On reception of the information node 5 picks a different color, and so on. This process continues until all the nodes have colored their edges. Then, edge orientation is applied to the edges with the same color. So, for instance, in Fig. 7(c) let us imagine the case where node 4 transmits to 6. Because of the node 4 transmission, the level of interference may be sufficiently high to corrupt the transmission of the link $(2,3)$



(a) Algorithm begins with node 6 (b) Edges' coloring (c) Edges' orientation

Fig. 7. Edge coloring algorithm

The same problem is reexamined by (Al-Khdour & Baroudi, 2010), under the assumption that nodes can communicate with different frequencies. Nowadays radio chips support multichannel transceivers which can help to reduce the number of required time slots in a TDMA frame. The distributed heuristic algorithm proposed in this work is based on solving the TDMA problem in a tree structure. The base station collects the information from its children to calculate how many slots are needed (e.g. 3 slots are required in Fig. 8(a)). Next, every parent allocates a time slot to its children 8(b). Each branch of the tree will use a different channel (the frequencies can be repeated in space), whereas the nodes in one branch will transmit in different slots.



(a) Frame's size calculation

(b) Distribute time slot allocation

Fig. 8. Time slot allocation based on a tree structure

In other versions of the problem the scheduling solution must satisfy certain requirements such as connectivity, data rates and delay. Kedad et al. (2010) formulate the problem as follows: construct a frame with the minimum number of time slots such that at each time slot the activated links are not in conflict, and form a strongly-connected graph. The second constraint ensures that a node will be able to send a data packet to any other node in the network through the activated links. The links will be in conflict if they have the same transmitting or receiving node, or if the transmitting node of one link is the same as the receiving node of the other link. They show that this problem is NP-hard and propose two approximation algorithms. In (Ergen & Varaiya, 2010) the problem is to find an available slot allocation with minimum frame length, taking into account the quantity of data that a node needs to transmit. Notice that a link can be scheduled more than once in a time frame to satisfy the node data rates, which is the main difference with the basic version. Wang et al. (2007) formulate a multi-objective optimization problem. The question is to find a time slot allocation that satisfies i) the data delivery delay and ii) the node energy constraint. Here, not only are the transmitting and receiving energies taken into account, but also the energy consumed in switching between sleep and active modes. The two selected objectives contradict each other, since the energy objective seeks to maximize the number of nodes that are turned off, which in turn increases the delay. The trade-off between energy and delay is solved using the particle swarm optimization approach. This example gives a meaningful illustration of interdependence between problems coming from different layers. We have here a scheduling problem combined with a routing one in the sense that the latter one is responsible for the delay.

While TDMA-based approaches schedule transmissions sequentially over time, FDMA-based approaches permit multiple concurrent transmissions between neighboring nodes by allocating different channels/frequencies to them. Sensors in the network can thus tune their operating frequency over different channels to avoid interference and packet collisions in the network. One of the advantages of FDMA is the improvement of network throughput and packet transfer delay. In FDMA the problem can also be modeled as a graph-coloring problem, given that no two adjacent nodes are allowed to use the same channel.

Yu et al. (2010) show that the problem of assigning the channels such that *the interference* is minimized is NP-hard. They model the problem as a game where every node is a player and the interference is the objective to be minimized. Their algorithm assumes that routing is based on the tree structure. In each iteration an intermediate node selects its own channel

so as to cause the least possible interference for its neighboring nodes. The interference is calculated using local data that include the number of interfering parents in different branches existing in their neighborhood, their respective numbers of children, and whether or not these children are leaves within the tree structure. Notice that neighbors of a given node can belong to different branches and have different roles, parents or leaves. Based on an empirical study, Wu et al. (2008) find it more appropriate for a WSN to communicate using a single channel, but they suggest harnessing channel diversity by spreading the frequencies in space. They therefore propose a node-disjoint tree structure where every branch (subtree) communicates via a given channel. The objective here is to divide the network into multiple disjoint subtrees such that the interference between them is minimized. They show that the problem is NP-hard and propose a greedy heuristic.

CDMA spreads the baseband signal using different *Pseudo Noise (PN) codes* to enable multiple concurrent transmissions. In WSN, a PN code may be implemented as an attribute in the packet header (nodes simply need to check whether the code in an incoming packet matches their own set of codes) in order to reduce the complexity of modulation and decoding techniques in comparison to CDMA implementations using other technologies. Optimization problems relating to code allocation in CDMA are slightly different from those relating to time or channel allocation. For instance, in CDMA it is possible that two neighboring nodes share the same code but only one can use the code for transmitting while the other node can use it for receiving. The optimization problem may require that no two adjacent directed links have the same code. The difference between WSN and other wireless CDMA networks is not really to be found in the problem of code allocation, but in the CDMA concept itself. CDMA codes are not completely orthogonal. The high density of sensors in the network makes the problem of interference in concurrent transmission a very serious one. High interference causes problem for receiver nodes because they cannot 'understand' the signal addressed to them. In the literature the pure CDMA problem is addressed simultaneously with the channel and slot allocation problems. The problem of *channel and code allocation* to reduce interference is discussed in (Liu et al., 2006). Their distributed solution is a heuristic which tries to solve first the problem of channel allocation and subsequently the code allocation one. When CDMA is combined with *scheduling*, Chen et al. (2006) looks for a feasible schedule for all the nodes in the network, together with their respective PN codes such that there is no interference (or the interference falls below a given threshold) in any time slot and the total energy consumption is minimized.

3.5 Duty cycle

The node duty cycle is determined by its activity and sleep periods. During the sleep periods the sensor nodes do not consume energy, and so short activity periods mean energy savings. However, this has to be scheduled, because nodes can communicate with each other only during the activity periods. The set of active nodes in the network at a given moment must satisfy certain requirements, the most important being connectivity and coverage. In the first paragraph below we discuss the problem of node scheduling with a connectivity constraint. In the second paragraph coverage is taken into account and two additional constraints are introduced, namely i) life dependency between sensors and ii) connectivity.

(Nieberg, 2006) models the node duty cycle with a connectivity constraint as the MCDS problem. He also proposes a distributed algorithm for finding this set of nodes. Here it is assumed that the network is very dense and nodes are close to each other such that a

large number of nodes can become passive while the remaining nodes continue to ensure a connected structure. The active nodes correspond precisely to the CDS. According to the algorithm some nodes will have a special role: those nodes that form a Maximal Independent Set perform the role of *anchors*, and nodes used to connect anchor nodes perform the role of *bridges*. Nieberg (2006) shows that the set of anchor and bridge nodes forms the CDS. The initialization phase has self-organizing properties. Each node will try to get an active time slot according to the TDMA scheme. Then, any other node that enters into the network needs to decide locally whether or not it will be active (either as a bridge or as an anchor). The decision is based on the information provided by the neighbor nodes. This information includes the neighbor node ID, a list of all time slots showing the slots occupied by them and their respective neighbors, their role as an active node, and some synchronization information. When a node observes that there are less than two anchor nodes in its neighborhood for a given time slot, it becomes an anchor otherwise it seeks for the existence of bridge nodes. If it finds that any pair of anchor nodes are connected with bridges, it becomes passive.

The node duty cycle is also related to coverage requirements. Because monitoring is one of the main objectives of a WSN, the active nodes have to guarantee that a set of given targets will be monitored throughout the lifetime of the WSN. The problem is to group the nodes such that i) each group (known as a cover) is able to cover the targets and ii) the groups form disjoint sets of nodes in order to maximize the WSN lifetime. Usually a redundant sensor network is considered in this case. This question has elements of both a coverage problem (targets which need to be covered) and a scheduling problem. Only the nodes belonging to a cover are to be activated, while the others are to be put to sleep, and the covers are to be activated in a sequential manner. Cardei & Du (2005) have shown that this problem is NP-hard. In (Rossi et al., 2010), the problem is modeled as a linear program whose aim is to maximize the sum of the different covers' lifetimes, the constraint being that the total duration of a nodes' activity periods does not exceed the lifetime of its battery. The problem is solved using the column generation method. Aioffi et al. (2007) model the problem as the weighted set cover problem (WSCP). Given n sets $(S_1, S_2, ...S_n)$ formed from elements of a universal set denoted the US, together with their associated activation costs, WSCP seeks to find a subset of these sets such that the sum of the activation costs is minimized, and whose union corresponds to the US. The set of the targets in the network problem is modeled as the US, the sets S_i represent the set of the target covered by sensor i, and the cost of S_i is the inverse of energy for the sensor i. The problem is solved off-line and the results are fed into the sink. When the mobile sink gathers data from the nodes, it also indicates to them whether they will be active in the following period. This method is used particularly for this case because the number of possible combinations is exponential (the number of constraints is very small) and it can achieve faster convergence.

(Dhawan & Prasad, 2009) remove the constraint of disjoint covers. If a node is included in two or more cover sets, then its energy capacity will influence the life of these sets. They therefore propose a solution based on the construction of a life dependency (LD) graph. In this graph covers are represented by vertices, linked by an edge if they share the same sensors. The LD graph is introduced into the problem in order to identify the covers having the least impact on the other covers. Their distributed approach adds a communication cost between neighboring nodes which need to exchange information such as the remaining energy and the region (area or targets) they can cover. A further cost is added, corresponding to the processing of this information and to making a decision. Every sensor thus needs to construct an LD graph

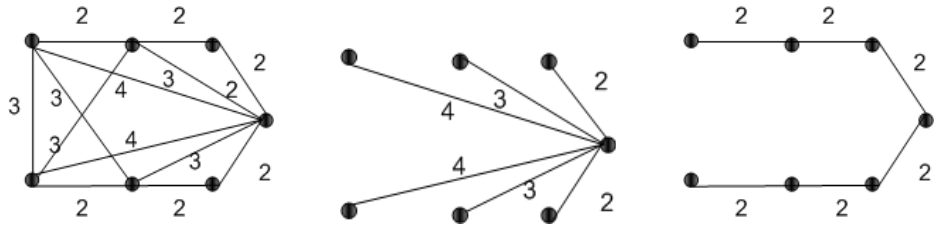
based on its local information and to identify the cover with the smallest impact in order to be part of it. Finally, there is also a communication cost corresponding to the negotiation phase where nodes attempt to obtain a stable solution. In (Cardei & Cardei, 2008; Zou & Chakrabarty, 2005) the same problem is discussed and an additional constraint imposed: each set is required to be connected with the base station. In (Cardei & Cardei, 2008) the problem is formulated as Integer Linear Programming. It is first centrally solved using ILOG CPLEX, and then via a distributed approach. In the distributed case each node needs to know not only its own coordinates but also those of the given targets and base station. The initialization phase has a considerable communication cost resulting from exchanging the list of targets that the two-hop neighbors cover, the status of every node, and the synchronization message. This initialization phase includes the creation of the cover sets, while the subsequent phase finds the relaying nodes for connecting the cover with the base station (one node in the cover constructs a spanning tree that includes the target set and the BS).

4. Routing

Data transmission in WSNs, also referred to as the routing problem, is one of the most widely studied problems in WSN. Different to the previous section, we focus here on the main proposed models and give some analysis on their use. The models and methods used for solving routing problems in WSN can be roughly divided in two main groups. The first group includes related shortest and spanning tree models, while the second group is centered around flow models and comprises a range of different minimum cost/maximum multicommodity flow models. While abundant work relating to such problems exists for wired networks, some new challenges have appeared for wireless networks, and especially for WSNs. The nature of some of these problems can change quite radically when they are placed in a WSN context and new requirements are introduced. These requirements include sensors' energy constraints, the interference caused by the broadcast nature of transmissions over wireless links, as well as data compression, aggregation and processing constraints. For instance, in traditional formulations of the network flow problem, link capacity is a strong constraint, while in WSN this constraint is frequently supplanted by the node energy constraint. Another important difference between these two paradigms is the inclusion of the dynamic topology models and the need for distributed solutions for wireless sensor networks.

4.1 Shortest Path and Spanning Tree based models

Shortest Path Tree (SPT) and Minimum Spanning Tree (MST) remain widely used models for routing design, even in WSNs. The goal of a SPT is to find a path of minimum cost from a specified *source node* to another specified *sink node*, assuming that each edge has an associated cost. In the WSN context the edge cost usually represents the power that would be consumed by the transmitting node when sending a packet to the node at the opposite end of the edge. Distributed routing algorithms based on Dijkstra, Bellman-Ford or Chandy-Misra's distributed algorithms can thus be employed (Rodoplu & H., 1999; Yilmaz & Erciyes, 2010). One of the disadvantages of SPT is the unbalanced load between the sensors and the disparity in the energy used by them that such methods can lead to. To overcome this problem, different strategies are proposed. In (Yilmaz & Erciyes, 2010) every node can regenerate a path when a fault occurs or available energy is depleted. Other works consider edge cost to be a combination of several metrics such as residual energy, buffer size, or the number of neighboring nodes.



(a) A simple graph (b) Short path tree (c) Minimum spanning tree

Fig. 9. Shortest path (b) and minimum spanning tree (c) for the graph shown in (a)

Going further, WSN brings new constraints which may modify the nature of the problem. For instance, many applications of WSN require that the intermediate or relay nodes aggregate the data, while the criterion used is minimizing energy consumption. For (Cristescu et al., 2006) the joint problem of data aggregation and routing is NP-hard, and their heuristic combines an MST with an SPT. Normally, in cases where there is a high aggregation coefficient, the amount of traffic increases slightly from the source to the sinks, and an MST is a good compromise. However, where the aggregation coefficient is low, routes need to be found that minimize the number of hops, and therefore an SPT should be constructed. MST is the tree structure which minimizes the sum of edge costs, and the problem is polynomial. The difference between a shortest path tree and a minimum spanning tree is shown in Fig. 9.

Minimizing the total energy consumption is, however, not enough, since some nodes deplete their energy faster than others and may cause network partition. To balance the energy consumption, one strategy is to minimize the maximum energy consumption of the nodes. This problem has been modeled by (Gagarin et al., 2009) as the minimum degree spanning tree (MDST), which is an NP-hard optimization problem. Variations of this problem are encountered in the literature, in (Erciyes et al., 2008; Huang et al., 2006). A joint routing and data aggregation problem is also discussed in (Karaki et al., 2009) for a two-tier network, and some heuristic algorithms such as GA and greedy are proposed. From a distributed perspective, adapted versions of Prim's and Kruskal's algorithms have been proposed in (Attarde et al., 2010). In the distributed versions of SPT a node need only communicate to its neighbors information concerning the cost of links. Each node decides to communicate with the node that provides the minimal cost to the base station. An ACK mechanism is needed to dictate the end of the process. It may be remarked here that almost all the above cited models lead to single path routing schemes. They have the great advantage of being simple from an implementation point of view, while their main drawback is their difficulty in embracing additional requirements, energy consumption in particular. We now present some flow-based models that can model such requirements in a suitable way.

4.2 Flow-based models

The need to include energy/capacity constraints leads naturally to the use of flow models. Particularly for the WSN, routing problems are formulated as MultiCommodity Flow Problems (MCFPs). The commodity is a source-destination pair, and we are faced with an MFCP whenever several commodities share the network resources. In an MCFP the commodities will have different sources and/or destinations, but they are bound together

insofar as they share the same link capacities. Regarding commodities, a WSN gives rise to either single-sink or multi-sink models, and in the case of single-sink models all commodities will have the same extremity, namely the base station. In the following subsection 4.2.1 we discuss some basic versions of flow models used for routing path calculation in WSN. Then, in subsection 4.2.2 some further extended routing problems are presented.

4.2.1 Conventional flow models in WSN

A standard flow problem in WSN (regardless of whether it is a multicommodity flow problem) includes two type of constraints, namely the flow conservation constraint and the energy constraint.

$$\sum_{j \in N_i} x_{ij}(t) = \sum_{j \in N_i} x_{ji}(t) + y_i(t) \qquad \forall i \in N, \forall j \in N_i, \forall t \in T, \tag{5}$$

$$\sum_{t \in T} \sum_{j \in N_i} x_{ij}(T) * e_{ij} \leq E_i \qquad \forall i \in N, \forall j \in N_i, \tag{6}$$

where t (respectively T) is a time instance (respectively the network lifetime), N the set of sensors, N_i the set of neighboring nodes of i, x_{ij} the flow over the edge ij (that is to say the data transmitted over this link), y_i the data generated by node i, e_{ij} the energy consumed in transmitting a unit flow and E_i the initial energy of the sensor. The flow conservation constraint, Equation (5), shows that the total amount of flow that a sensor receives plus the amount of data that it generates is equal to the amount of information that it transmits. The second constraint given in Equation (6) is the capacity constraint, which is related to energy. This constraint implies that the energy consumed by a sensor for transmitting the flow throughout the lifetime of the network must be less than its initial energy. In standard network flow problems this constraint is usually related to link capacity.

One of the first works to formulate this problem in terms of Integer Linear Programming is to be found in (Chang & Tassiulas, 2004). The flow is represented here by the number of packets and the transmission energy is calculated based on the distance between the nodes (and hence assuming a power control mechanism). The optimal solution of this problem gives an upper bound for network lifetime. While the problem of lifetime or flow maximization under these constraints can be solved in polynomial time for continuous values of flow x, the integer version is shown to be strongly NP-hard in Bodlaender et al. (2010). The distributed version of this problem is discussed in (Madan & Lall, 2006), where the subgradient algorithm is used to solve the problem. At each iteration the algorithm estimates the gradient value at a given point of the objective function and determines the next point to be considered, until the optimum is reached. The distributed implementation of this algorithm requires that every node keeps track of two variables, namely the *flow rate* of every outgoing link and the *network lifetime*. These variables are updated during each iteration of the algorithm based on their previous values and the subgradient function values (also a function of flow rates and network lifetime) are calculated according the information received from neighbor nodes. Subgradient methods are also used by Rabbat & Nowak (2004) as convenient tools for designing a distributed approach in sensor networks. Another characteristic of WSNs is the data aggregation applied by nodes. This phenomenon can easily be taken into account by slightly modifying the conservation flow constraint. For instance, in Cheng et al. (2009) each node sends the maximum amount of information between the received and the generated data set as in Equation (7).

$$\sum_{j \in N_i} x_{ij} = max_{j \in N_i} \left\{ x_{ji}, y_i \right\} \qquad \forall i \in N \tag{7}$$

The routing problem with data aggregation for lifetime maximization in a network has been formulated by Xue et al. (2005) as a concurrent multicommodity flow problem. Here the flow constraint implies that the amount of the flow commodities transmitted from a sensor node cannot be less than the sensor's data. They propose a polynomial time approximation scheme, strongly inspired by the Garg-Konemann algorithm. In outline, their algorithm is as follows: construct the shortest path between every source and the sink, initialize a cost unit flow for every node, push the maximum possible flow along the path for every commodity, update the cost of energy for every node and repeat the process.

As regards routing paths, the routing schemes can use several paths (in other words perform multipath routing), or a single path (single-path routing.) Although requiring routing via a single path would appear preferable for WSN, adding such a constraint to the mathematical model gives rise to NP-hard problems. Worth citing here are two approaches proposed for WSN that attempt to circumvent the computational burden of such models while providing simplicity in implementation. The first approach computes a solution involving multiple paths, but uses only one single path at a time. Hou et al. (2004) propose an algorithm to solve the problem in two phases. In the first phase a solution is found for the multipath routing problem. Consequently every node knows the set of the relaying nodes and the respective amount of information to send to them. In the second phase one node, according to some local rule, will select one of its relaying nodes and will transmit to it the whole amount of information to be sent in this round. The second approach, in stark contrast to the first approach just described where routing takes place from the sensors to the BS (i.e. flat routing), may be seen as hierarchical routing, in that it decomposes the data transmission into two levels and thus converges to a cluster-based scheme. Each cluster head (CH) receives the data from the nodes of its cluster and from the other CHs, and transmits this data to another CH in the direction of the BS. Bari et al. (2008) consider a two-tier heterogeneous network containing powerful relay nodes which form a connected network that can relay data to the BS. They formulate the optimization problem as follows: knowing the positions of sensors and relay (CH) nodes, how should the network be clustered in order to maximize its lifetime? A sensor is not obliged to transmit directly to the CH, and sensors may have different amounts of flow to transmit. The problem is formulated as a max-min LP. Because the decision variables can take only binary values (1 if the sensor belongs to a given cluster and 0 otherwise) and the flow rate variable corresponds to a number of bits, we are dealing with an ILP problem. The heuristics presented for this problem are centralized. Other centralized techniques for solving the clustering problem in WSN are based on Fuzzy Logic (FL) (Anno et al., 2007; Ran et al., 2010), while Mehrjoo et al. (2011) proposes genetic algorithms.

4.2.2 Enhanced flow based models

Advances in technology and the broad range of applications for WSN have given rise to new QoS requirements and made routing a more complex matter. Interference, delay and questions of reliability may all place additional constraints and lead to more elaborate and challenging versions of routing problems. All this will be in the focus of this paragraph.

Radio interference has a significant impact on the performance of WSN as it affects the functioning of both MAC and routing protocols, and directly affects the transmission capacity

of links. In contrast to traditional networks where the capacity of links is determined by physical parameters only, in wireless communications radio interference strongly affects the transmission capacity of links that are located close to one another. The models we have cited above assume that the quantity of information generated is sufficiently low, or the channel capacity sufficiently high, for transmission capacity not to be an issue. But this assumption clearly does not always hold, and capacity constraints over links are sometimes unavoidable. It should be noted that IEEE 802.15.4 defines data rates of 20, 40, or 250 Kb/s for the physical layers. Channel capacity may therefore represent a strong constraint where huge amounts of data need to be transmitted, or when many sources have to transmit simultaneously. Interference needs to be taken into account because of the high bit error rates that it may cause. The capacity of wireless channels is calculated from the Shannon-Hartley formula given in Equation (8).

$$C = B \cdot \log_2 \left(1 + \frac{S}{N}\right) \tag{8}$$

where C is the channel capacity (in bits per second), B the channel bandwidth (Hz) and S/N the signal-to-noise ratio.

From the point of view of computational complexity, including this constraint in the model makes the problem NP-hard, as shown in (Jain et al., 2003). More precisely, they show that the problem of finding a maximal flow for a source-destination pair under the interference constraint is equivalent to the Minimum Independent Set problem in a graph, and therefore NP-hard.

Krishnamachari & Ordonez (2003) add the link capacity constraint to the basic version of the flow problem with the goal of maximizing the throughput or minimizing the overall energy consumption. To ensure that the solution will not generate scenarios in which the traffic load is unfair for the nodes in the network, the flow transmitted by a node has to be less than a given fraction of the total flow generated by the network. Patel et al. (2006) add the following two constraints to the basic version of the routing problem: (i) the link capacity constraint where the rate (the number of packets per unit time) at each link has to be smaller than its capacity, and (ii) the node capacity constraint where the number of packets that a node can process in a unit time has to be smaller than its given capacity. The proposed algorithm is centralized and aims to find a maximum flow with the smallest possible energy cost. It is a kind of combination of maximum flow (getting as much flow as possible from the source to the sink) and shortest path (traveling from the source to the sink with minimum cost). The problem addressed in (Xu et al., 2008) has the same structure as that found in Patel et al. (2006), but the objective is utility maximization, which is a nonlinear convex function of the transmission rate. The problem is solved using the Lagrangian method. This method attempts to decompose the problem into a number of sub-problems via a Lagrange multiplier and to solve each of them separately. In these problems it is assumed that the bandwidth B is shared between different node channels, or that the nodes use the whole bandwidth but are already scheduled in order to avoid interference.

There are two possible ways of modeling a successful transmission in the presence of interference: i) the physical context, which requires that the Signal-to-Interference and Noise Ratio (SINR) given in Equation (10) exceeds a certain threshold; ii) the protocol context, where no two neighboring nodes may transmit at the same time.

Routing under the physical interference model is more complex. Wang et al. (2011) discuss a link scheduling problem where flow capacities are satisfied and the time taken for scheduling is minimized. In this case the channel capacity is variable over time due to SINR, and its integral gives the service provided by the channel as expressed in Equation (9).

$$C_{ij}(t) = \int_0^t B \cdot log(1 + SINR_{ij}(\tau))d\tau \tag{9}$$

where $C_{ij}(t)$ is the channel service of link (i, j) during time t, and B is the channel bandwidth.

$$SINR_{ij} = \frac{\omega_{ij}(t)P_i}{\sum_{k \in V^+/\{i\}}(\omega_{kj}(t)P_k) + N_a} \tag{10}$$

where $SINR_{ij}$ is the $SINR$ parameter for the link (i, j), ω_{ij} the gain of the fading channel for the link ij, P_i the power transmission of node i, $\omega_{kj}P_k$ measures the interference of the other links over the link (ij) and N_a is the floor noise which is a constant. The channel service calculated in each time slot is used as parameter to bound the link data rate. The problem is solved off-line using the column generation method.

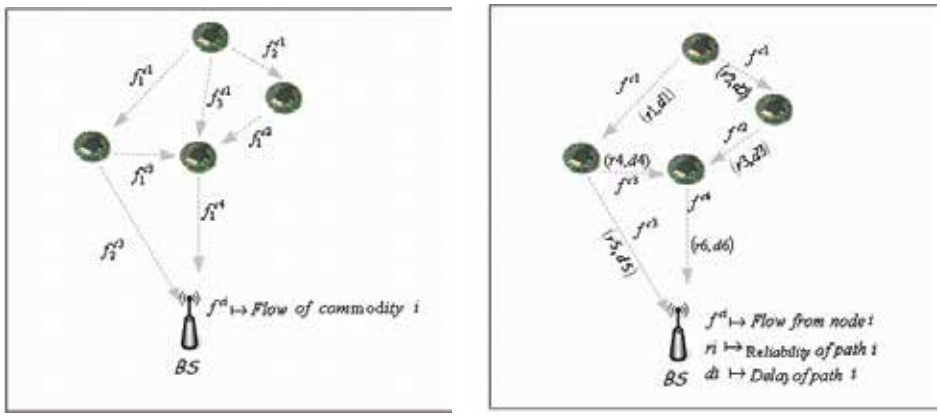
Interference can be more easily modeled in a protocol context. Wang et al. (2008) study the routing problem in the presence of interference by scheduling the nodes in accordance with the TDMA approach. The constraint added for the interference implies that the sum of the number of times a link is scheduled plus the sum of the number of times that all the links in its interference zone are scheduled in the time frame has to be smaller than the frame size, as in Equation (11).

$$N(e) + \sum_{e' \in I(e)} N(e') \leq S \tag{11}$$

where $N(e)$ is the number of times that the edge e is scheduled in the time frame, $I(e)$ is the subset of links of the original graph that can be influenced from e transmissions and S is the number of time slots in the frame.

We shall now focus on how WSN takes some QoS requirements and their associated metrics into consideration. We begin with a discussion of QoS metrics and the computational complexity that they introduce. Different metrics have different composition rules. Metrics such as delay, delay jitter and cost are additive (an additive metric is a metric which obeys the additive rule, meaning that the path metric is equal to the sum of the metric links that compose the relevant path). A multiplicative metric is a metric which obeys the multiplicative rule, meaning that the path metric is equal to the product of the link metric for all the links that compose the relevant path. Metrics like reliability (the probability that the transmission was successful) can thus be seen to be multiplicative. Finally, concave metrics obey the concave rule, meaning that the path metric is equal to the minimum (or maximum) link metric for all the links that compose the relevant path. Bandwidth is an example of a concave metric. Fig. 10 illustrates the concept of multicommodity flows in a graph and QoS multipath routing with two metrics.

In (Wang & Crowcroft, 1996) it is shown that the problem of finding a path which satisfies N additive metrics, and/or K multiplicative metrics (where N and K are positive integers) is NP-hard, while it becomes polynomial when one is concave and the other additive or multiplicative.



(a) multicommodity flows in a network (b) QoS multipath routing with two metrics

Fig. 10. Multicommodity and multipath Routing

Most works dealing with QoS routing in WSN are concerned either with finding (disjoint) paths for guaranteeing network resilience (fault-tolerant network), or with finding a minimal number of paths such that QoS requirements are met. We recall that the problem of finding k disjoint paths (edge or vertex disjoint) such that the total cost of the paths is minimized has been shown in (Li et al., 1992) to be NP-hard, even for $k = 2$ in directed graphs. Heuristics therefore provide practical approaches for solving these kinds of problems. (Okdem & Karaboga, 2009) report an approach combining ACO with a tabu search. Each source node wishing to transmit data toward the BS has to launch n ants (n corresponds to the number of data packages that the source transmits). The ant's movement is based on the probabilistic decision where the heuristic value represents the estimation of the residual energy. After all the ants have completed their journey (from source to destination), each ant k deposits a quantity of pheromone equal to the inverse of the total number of nodes included in the path. This task is performed by sending ant k back to its source node following the arrival path. In this type of ACO each receiver node has to maintain a tabu list with the identities of the ants that it has encountered, enabling it to decide whether to accept the upcoming packet of ant k. Routing the information efficiently to guarantee the delay and reliability constraint is discussed in Saleem et al. (2010), who proposes a multi-agent approach for ant colony optimization (ACO). The movement of the ant is guided by the probabilistic decision rule, equation (4). The pheromone value corresponds to the end-to-end delay. The two heuristic evaluation parameters of every edge are determined by the residual energy at the extremity of the edge and its packet receive rate (PRR).

In (Bagula & Mazandu, 2008) the QoS routing problem is concerned with delay and reliability criteria. The goal is to find the smallest set of disjoint paths between a source and a destination such that both criteria are satisfied and energy consumption is minimized. Delay is a stringent metric, meaning that if the delay is not respected in any of the set paths, the packet is dropped. In contrast, the reliability of every source-destination connection obeys the multiplicative composition rule. Hence the more paths in the set, the more reliable the set will be. The problems of finding the path which minimizes the energy or the delay, or maximizes the reliability, taken separately, are solvable in polynomial time, but the problem considered in its entirety is not.

5. Open issues and concluding remarks

There are several issues in WSN which are still open or which have not been sufficiently addressed.

- Dynamicity is one of the most noticeable characteristics of WSN and also one of the biggest challenges. The term covers such phenomena as node failure, link fluctuations, node attacks and mobile nodes. Many studies in routing, coverage, scheduling or topology control have attempted to find solutions where these events occur, but including them in optimization problem models remains a challenge.
- We consider that scalability is an important issue which is frequently neglected when solution methods are proposed. The eventually changes in network dimensioning may sometimes require to resolve the problem or to sufficiently increase the computation time. We observe this particularly in relation to issues related to multi-sink/multicommodity design and network cross layer design.
- With respect to coverage problems, there are several potential directions that have not been fully explored. These include solving the deployment problem in the presence of obstacles, taking into account the restrictions for node placement and $3D$ deployments. In routing and topology control, cooperative decision-making strategies and opportunistic approaches also need to be modeled and examined in optimization problems, since in both areas some of the problems discussed here have been successfully addressed through opportunistic approach. But not many theoretical works have been undertaken in relation to this paradigm. Many questions remain open. For instance, in what scenarios should an opportunistic approach be favored over other approaches? How close is an opportunistic approach solution likely to be to the optimal solution? Routing in opportunistic networks [1] adopts a people-centric approach to model the network semantics (Verdone & Fabri, 2010). This routing group is classified as sociability-based routing and has been modeled in (Yoneki et al., 2007) based on human behavior characteristics. They propose a Socio-Aware Overlay (multi-point event dissemination using an overlay constructed by closeness centrality nodes in communities) for publish/subscribe communication. It is not clear whether these strategies might be appropriate for WSN.
- Another crucial issue is the difference that still exists between theoretical studies and practical implementations in WSN. Some theoretical studies have already presented models for cross-layer design, together with corresponding solutions. But many of them remain centralized and require off-line computation. We remark that in some mathematical formulations the variables are considered continuous, despite the discontinuous nature of the corresponding events such as power transmission and flow. On the other hand, algorithms or protocols implemented in real hardware or tested in simulations do not address cross-layer design. They aim at distributed and on-line computations and handle mostly simplified problems. Moreover, in these works the analyses that might yield an optimal solution are neglected, and it is difficult to grasp the problem complexity and to know whether there is room for further improvement. Combining these two approaches is far from straightforward and calls for substantial work. We see as a primary concern in

[1] Examples of opportunistic networks are Delay Tolerant Networks (DTN) (Pelusi et al., 2006.) or Pocket Switched Networks, VANETs, networks composed of devices such as MP3 players, mobile telephones and PDAs which can communicate with each other by Bluetooth or Wi-Fi to share data, or even wireless sensor networks which can send data using technologies such as GSM/UMTS, WiFi, etc.

this context the development of optimization tools and dedicated software to bridge the gap between optimization methods and their practical implementation in WSN.

- Finally, we consider that uncertainty has received very little attention until now. Nonetheless, uncertainty is an important characteristic inherent in the nature of WSNs, and is related to different aspects such as event detection, sensor location and data delivery. Some attempts to model these situations use probabilities associated with these different kinds of events. The main difficulties in taking the uncertainty of WSNs into account are twofold. First, measuring the distribution of events is not an easy task and is both environment- and application-dependent. Secondly, despite recent advances in robust optimization[2] tackling probabilistic optimization problems is not for the faint-hearted.

To conclude, wireless sensor networks represent an attractive research area due to several factors as the resource-constrained nature of sensor nodes, interference, data aggregation, power consumption model and the wide range of both commercial and military applications that this technology offers. Successful network design and deployment include understanding and modeling several problems related to these factors, which ultimately determine the available range and data rate of a WSN, as well as cost and battery lifetime. Therefore this study, intended to researchers and graduate students in computer science and fields related to operations research, information technology and applied mathematics, gives some highlights on a number of representative network problems in WSN and focuses on their respective optimization problems.

6. References

Aioffi, W., Mateus, G. & Quintao, F. (2007). Optimization issues and algorithms for WSNs with mobile sink, *Intern. Netw. Opt. Conf.* pp. 1–6.

Al-Khdour, T. & Baroudi, U. (2010). An energy-efficient distributed schedule-based communication protocol for periodic wireless sensor networks, *Arab. jour. for sci. and eng.* 35: 155–168.

Anno, J., Barolli, L., Xhafa, F. & Durresi, A. (2007). A cluster head selection method for wireless sensor networks based on fuzzy logic, *IEEE TENCON* pp. 1–4.

Attarde, S. A., Ragha, L. L. & Dhamal, S. K. (2010). An enhanced spanning tree topology for wireless sensor networks, *Int. Journal of Comp App.* 1: 46–51.

Averbakh, I. & Berman, O. (1997). Minimax regretp-center location on a network with demand uncertainty, *Elsevier Science* 5: 247–254.

Bagula, A. B. & Mazandu, K. G. (2008). Energy constrained multipath routing in wireless sensor networks, *Proc. of Ubiquitous Intell. and Comp.* pp. 453–467.

Bari, A., Jaekel, A. & Bandyopadhyay, S. (2008). Clustering strategies for improving the lifetime of two-tiered sensor networks, *Comp. Comm.* 31: 3451–3459.

Bellman, R. (1957). *Dynamic Programming*, Princeton University Press.

Bertsimas, D. & Sim, M. (2004). The price of robustness, *Operations Research,* 1: 35–53.

Bodlaender, H., Tan, R. B., van Dijk, T. & van Leeuwen, J. (2010). Integer maximum flow in wireless sensor networks with energy constraint, *Technical report*, Utrecht University.

Cardei, I. & Cardei, M. (2008). Energy-efficient connected-coverage in wireless sensor networks, *International Journal of Sensor Networks* 3: 201–210.

[2] Following the works of Bertsimas & Sim (2004), who showed how to model a stochastic optimization problem as a Linear Program under weak conditions, robust optimization has been intensively investigated.

Cardei, M. & Du, D. (2005). Improving wireless sensor network lifetime through power aware organization, *Wireless Networks* 11: 333–340.

Cavalier, T. M., Conner, W., Castillo, E. & Brown, S. (2007). A heuristic algorithm for minimax sensor location in the plane, *Europian Journal of Operational Research* pp. 42–55.

Chang, H. & Tassiulas, L. (2004). Maximum lifetime routing in wireless sensor networks, *IEEE trans. on Netw.* 12: 609–619.

Chen, M., Oh, C. & Yener, A. (2006). Efficient scheduling for delay constrained multi-rate CDMA systems, *Spread Spectrum Techniques and Applications* pp. 371–375.

Cheng, M., Gong, X. & Cai, L. (2009). Joint routing and link rate allocation under bandwidth and energy constraints in sensor networks, *IEEE Trans. on Wir. Comm.* 8: 3770 – 3779.

Cheng, X., Narahari, B., Simha, R., Cheng, M. X. & Liu, D. (2003). Strong minimum energy topology in WSNs: NP-Completeness and heuristics, *IEEE Trans. on Mob. Comp.* 2: 248 – 256.

Cristescu, R., Lozano, B., Vetterli, M. & Wattenhofer, R. (2006). Network correlated data gathering with explicit communication: NP-Completeness and algorithms, *Networking, IEEE/ACM* 14: 41–54.

Dantzig, G. B. (1963). *Linear Programming and Extensions*, Princeton.

Dhawan, A. & Prasad, S. K. (2009). A distributed algorithmic framework for coverage problems in wireless sensor networks, *Proc. of Parallel and Distrib. Proc.* pp. 18–25.

Dorigo, M., Maniezzo, V. & Colorni, A. (1996). The ant system: Optimization by a colony of cooperating agents, *IEEE Trans. on System Man, and Cybernetics-Part B* 26: 29–41.

Duttagupta, A., Bishnu, A. & Sengupta, l. (2008). Maximal breach in WSNs: Geometric Characterization and Algorithms, *Algosensors* pp. 126 –137.

Efrat, A., Peled, S. & Mitchel, J. (2005). Approximation algorithms for two optimal location problems in sensor networks, *Broadband Networks* 1: 714–723.

Erciyes, K., Ozsoyeller, D. & Dagdeviren, O. (2008). Distributed algorithms to form cluster based spanning trees in wireless sensor networks, *Proc. of Computer Science* pp. 519–528.

Ergen, S. & Varaiya, P. (2010). TDMA scheduling algorithms for WSN, *Wireless Networks* 16: 985 – 997.

Fidanova, S., Marinov, P. & Alba, E. (2010). ACO for optimal sensor layout, *Proceeding of International Conference on Evaluationary Computation* pp. 5–9.

Gagarin, A., Hussain, S. & T., Y. L. (2009). Distributed search for balanced energy consumption spanning trees in wireless sensor networks, *Adv. Inf. Net. and App. Work.* pp. 975–982.

Gandham, S., Dawande, M. & Prakash, R. (2005). Link scheduling in sensor networks: distributed edge coloring revisited, *Infocom* 4: 2492 – 2501.

Glover, F. (1989). Tabu search - part i, *ORSA Journal on Computing* 1: 190–206.

Gogu, A., Nace, D. & Challal, Y. (2010). A note on joint optimal transmission range assignment and deployment for wireless sensor networks, *IEEE Networks* pp. 1–6.

Holland, J. (1975). *Adaptation in natural and artificial systems*, University of Michigan Press.

Hou, Y., Shi, Y., Pan, J. & Midkiff, S. (2004). Lifetime-optimal data routing inwireless sensor networks without flow splitting, *Workshop on Broadband Advanced Sensor Networks* .

Huang, G., Li, X. & He, J. (2006). Dynamic minimal spanning tree routing protocol for large wireless sensor networks, *IProc. of Indust. Electr. and Applic.* pp. 1–5.

Jain, K., Padhye, J., Padmanabhan, V. & Qiu, L. (2003). Impact of interference on multi-hop wireless network performance, *MobiCom, ACM* pp. 66 – 80.

Karaki, J., Ul-Mustafa, R. & Kamal, A. (2009). Data aggregation and routing in WSN : Optimal and heuristic algorithms, *Computer networks* pp. 945–960.

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming, *Combinatorica* 4: 373–395.

Kawano, R. & Miyazaki, T. (2009). Distributed data aggregation in multi-sink sensor networks using a graph coloring algorithm, *Proc. of Adv. Inf. Netw. and Applic.* pp. 906–912.

Ke, W., Liqiang, W., Shiyu, C. & Song, Q. (2009). An energy-saving algorithm of WSN based on Gabriel graph, *Wir. Comm., Netw. and Mob. Comp.* pp. 1–4.

Kedad, S., Pasqual, F. & Fouilhoux, P. (2010). Ordonnancement de paquets dans les réseaux sans fil, *In Proc. of ROADEF* .

Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization, *Proc. of IEEE Int. Conf. on Neural Netw.* 4: 1942 – 1948.

Konstantinidis, A., Yang, K., Chen, H. & Zhang, Q. (2007). Energy-aware topology control for WSN using memetic algorithms, *Computer Communications* pp. 2573–2764.

Krishnamachari, B. & Ordonez, F. (2003). Analysis of energy-efficient, fair routing in wireless sensor networks through non-linear optimization, *IEEE Vehicular Technology Conference* 5.

Kuhn, H. W.; Tucker, A. W. (1951). Nonlinear programming, *Proc. of 2nd Berkeley Symposium.* pp. 481–492.

Li, C., McCormick, S. & Simchi-Levi, D. (1992). Finding disjoint paths with different path-costs: Complexity and algorithms, *Networks* 22: 653–667.

Li, J. (2008). *Optimization Problems in Wireless Sensor and Passive Optical Networks.*, PhD thesis, The University of Melbourne, Australia.

Li, X., Calinescu, Y. & Wan, G. (2002). Distributed construction of a planar spanner and routing for ad hoc wireless networks., *In: Proc. of IEEE Infocom* pp. 1268–1277.

Liu, B., Otis, B., Chou, C. & Jha, S. (2006). A novel multi-channel CDMA system for wireless sensor networks, *Sensor Networks, ACM* 5: 1–30.

Liu, Y., Zhang, Q. & Ni, M. (2010). Opportunity-based topology control in wireless sensor networks, *Parallel and Distributed Systems* pp. 405–416.

Ma, J., Chen, Q., Qian, Z. & Ni, L. (2008). Opportunistic transmission based QoS topology control in wireless sensor network, *Mob. Ad Hoc and Sen. Sys.* pp. 422–427.

Madan, R. & Lall, S. (2006). Distributed algorithms for maximum lifetime routing in wireless sensor networks, *IEEE Transactions on Wireless Communications* 5: 2185 – 2193.

Meguerdichian, S., Koushanfar, F., Potkonjak, M. & Srivastava, M. B. (2001). Coverage problems in wireless ad hoc sensor networks, *Proc. of IEEE Infocom* .

Mehrjoo, S., Aghaee, H. & Karimi, H. (2011). A novel hybrid GA-ABC based energy efficient clustering in wireless sensor network, *Multimedia and Wireless Networks* 2: 40–45.

Moscato, P. (1999). *Memetic algorithms: a short introduction*, McGraw-Hill.

Nieberg, T. (2006). *Independent and dominating sets in wireless communication graphs*, PhD thesis, Twente University, Netherlands.

Okdem, S. & Karaboga, D. (2009). Routing in WSN using an ant colony optimization (ACO) router chip, *Sensors* pp. 909–921.

Patel, M., Chandrasekaran, R. & Venkatesan, S. (2006). Energy-efficient capacity-constrained routing in wireless sensor networks, *Int. J. Perv. Comp. and Comm.* pp. 69–80.

Pelusi, L., Passarella, A. & Conti, M. (2006.). Opportunistic networking: data forwarding in disconnected mobile ad hoc networks, *Communications Magazine, IEEE* .

Rabbat, M. & Nowak, R. (2004). Distributed optimization in sensor networks, *IPSN* pp. 20–27.

Ran, G., Zhang, H. & Gong, S. (2010). Improving on leach protocol of wireless sensor networks using fuzzy logic, *Journal of Information & Computational Science* 7: 767–775.

Ren, H., Meng, M. Q. & Chen, X. (2006). Investigating network optimization approaches in wireless sensor networks, *in Proc. of IROS* pp. 2015–2021.

Rodoplu, V. & H., M. T. (1999). Minimum energy mobile wireless networks, *Journal on selected areas in communications* 17: 1333–1344.

Rossi, A., Singh, A. & Sevaux, M. (2010). Génération de colonnes dans le réseaux de capteurs sans fil, *In Proc. of ROADEF* .

Rotar, C., Risteiu, M., Ileana, I. & Hutanu, C. (2009). Optimal sensors network layout using evoutionary algorithms, *Proc. of Inter. Conf. on Automation & information* pp. 88–93.

Saleem, K., Fisal, N., Baharudin, M., Hafizah, S., Kamilah, S. & Rashid, R. (2010). Colony inspired self-optimized routing protocol based on cross layer architecture for WSN, *Int. Conf. on Communications* pp. 178–183.

Shor, N. Z. (1985). *Minimization Methods for Non-differentiable Functions*, Springer-Verlag.

Sridharan, A. & Krishnamachari, B. (2004). Max-min fair collision-free scheduling for wireless sensor networks, *Perfor., Comp. and Communic.* pp. 585–590.

Suomela, J. (2009). *Optimisation Problems in Wireless Sensor Networks: Local Algorithms and Local Graphs*, PhD thesis, University of Helsinki, Finland.

Tao, W., Chen, C., Yang, B. & Guan, X. (2010). Adaptive topology control for throughput optimization in wireless sensor networks, *Proc. of Int. Conf. Comm. Technology* pp. 1299 – 1302.

Valli, R. & Dananjayan, P. (2008). Utility enhancement by power control in WSN with different topologies using game theoretic approach, *ICCIT* 2011: 85–89.

Verdone, R. & Fabri, F. (2010). Sociability based routing for environmental opportunistic networks, *Advances in Electr. and telecom.* 1: 98–103.

Wan, P., X., L. & Wang, Y. (2001). Power efficient and sparse spanner for wireless ad hoc networks, *In IEEE Int. Conf. on Comp. Com.Net.* pp. 564 – 567.

Wang, Q., Fan, P., Wu, D. & Ben Letaief, K. (2011). End-to-end delay constrained routing and scheduling for wireless sensor networks, *ICC* pp. 1–5.

Wang, T., Wu, Z. & Mao, J. (2007). A new method for multi-objective TDMA scheduling in WSN using pareto-based PSO and fuzzy comprehensive judgement, *Proc. of High Perf. Comp. and Comm.* pp. 144–155.

Wang, Y., Wang, W., Li, X. & Song, W. (2008). Interference-aware joint routing and TDMA link scheduling for static wireless networks, *IEEE Trans. on par. and distrib. sys.* 19: 1709–1725.

Wang, Z. & Crowcroft, J. (1996). Quality of service routing for supporting multimedia applications,, *IEEE J. Sel. Areas Commun* 14: 1228–1234.

Wu, Y. & Li, Y. (2008). Construction algorithms for k-connected m-dominating sets in WSN, *MobiHoc* pp. 83–90.

Wu, Y., Stankovic, J. A., He, T., Lu, J. & Lin, S. (2008). Realistic and efficient multi-channel communications in wireless sensor networks, *INFOCOM* pp. 1193–1201.

Xu, W., Chen, J., Zhang, Y., Xiao, Y. & Sun, Y. (2008). Optimal rate routing in wireless sensor networks with guaranteed lifetime, *IEEE Globcom* pp. 1–5.

Xue, Y., Cui, Y. & Nahrstedt, K. (2005). Maximizing lifetime for data aggregation in wireless sensor networks, *Mobile Networks and Applications* 6: 853 – 864.

Yang, H. & Cai, W. (2008). Distributed power control algorithm with multi-QoS constraints for wireless sensor networks, *Proc. of Int. Conf on Netw., Sens. and Contr.* pp. 1031–1036.

Ye, W. & Heidemann, J. (2003). Medium access control in wireless sensor networks, *Technical report*, ISI-TR-580, Information Sciences Institute.

Yilmaz, O. & Erciyes, K. (2010). Distributed weighted node shortest path routing for wireless sensor networks, *Communications in Computer and Information Science*, 84: 304–314.

Yoneki, E., Hui, P., Chan, S. & Crowcroft, J. (2007). A socio-aware overlay for publish/subscribe communication in delay tolerant networks, *Proc. of MSWiM, ACM* pp. 225–234.

Yu, Q., Chenyz, J., Fanz, Y., Shenz, X. & Suny, Y. (2010). Multi-channel assignment in wireless sensor networks: A game theoretic approach, *INFOCOM* pp. 1127–1135.

Yuanyuan, Z., Jia, X. & Yanxiang, H. (2006). Energy efficient distributed connected dominating sets construction in WSN., *Proc. of Wireless communications* pp. 797–802.

Zou, Y. & Chakrabarty, K. A. (2005). A distributed coverage and connectivity centric technique for selecting active nodes in wireless sensor networks., *IEEE Trans. on Comp.* 54: 978–991.

Part 4

Telecommunications

8

Telecommunications Service Domain Ontology: Semantic Interoperation Foundation of Intelligent Integrated Services

Xiuquan Qiao, Xiaofeng Li and Junliang Chen
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
China

1. Introduction

Network is the bearer of services and services are the soul of network. The convergent network extends the original communications service type and gradually forms new convergent services which integrate the traditional telecommunication services and a large number of value-added services or contents on Internet (Kolberg et al., 2010). The integrated service is essentially to handle the data and services across heterogeneous networks and service platforms. Facing the heterogeneity and diversity of service resources, integrated services need to run in a multi-terminal, multi-access network and multi-platform heterogeneous environment. These tremendous changes of service environment present a significant interoperability challenge for traditional service provisioning theory. Nowadays, the provision of context-awareness, adaptive personalized services is the development goal of future ubiquitous network (Park et al., 2009). It can enable seamless information exchange between humans, with humans and with entities (e.g., mobile devices), as well as entities and entities at any time, any place and in any way. To meet the development needs of adaptive personalized convergent services, dynamic service discovery and composition technologies are explored widely in the telecommunication service field (Bashah et al., 2010; Niazi & Mahmoud, 2009).

Today, semantic web service (McIlraith, 2001), as an establishing research paradigm, is defined as an augmentation of web service with semantic annotation, to facilitate the higher automation of service discovery, composition, invocation and monitoring in an open environment. Integration of the semantic web technology and telecommunications systems is explored widely in the telecommunication service field (Do & Jorstad, 2005; Vitvar & Viskova, 2005; Qiao et al., 2008a; Gutheim, 2011; Khan et al., 2011; Zander & Schandl, 2011). It is well known that ontology is the semantic interoperability and knowledge sharing foundation for semantic web services matching and context reasoning. Therefore, how to construct the telecommunications service domain ontology is an important factor of successfully applying semantic web services into telecommunication service systems (Veijalainen, 2007, 2008). However, telecommunication service field consists of a large number of concepts/terminologies and relations. How to abstract the sharing domain

concepts and reasonably organize them is a big challenge. Some related work has been done mainly in applying ontology technology to the mobile service domain. Based on the need for a standardized ontology that describes semantic models of the domains relevant for scalable NGN (Next Generation Network) service delivery platforms, the (Villalonga et al., 2009; Su et al., 2009) provide an overview of Mobile Ontology which comprises a core ontology and several subontologies, and its application examples in the service delivery platform. This work, as a part of IST SPICE project (IST SPICE project, 2008), is a meaningful attempt to establish a standardized ontology for mobile service delivery in NGN. In addition, IST SIMS project explored the semantic interfaces as a new means to specify and design service components and to guarantee compatibility in static and dynamic component compositions. And they also defined a domain-specific ontology, and its main purpose of the ontology is to establish a common description of the SIMS-related concepts and their semantics (Rój, 2008). The (Zhu et al., 2010) introduces a mobile ontology construction and retrieval system architecture. However, there lacks a general domain ontology modelling methodology for telecommunications service and the corresponding engineering approach to support the development work for domain ontology. The (Li et al., 2010) briefly introduced the constructing method of telecommunications service domain ontology (TSDO) proposed by our research team. However, the approach is not perfect at that time and still needs to be further improved. In fact, telecommunication service domain ontology, as the important semantic interoperability foundation of telecom network, still has no significant progress up to now. This has become the biggest obstacle to hamper the applications of semantic web technology in telecom field.

In this chapter, we clearly presented a practical domain ontology modelling approach for telecommunications service field. Under the guidance of this approach, our research team has created an open telecommunications service domain ontology knowledge repository which consists of around 430 telecommunications services-related ontology concepts/terminologies and 245 properties. Based on this domain ontology, we described the telecom network capability services in the semantic level to validate its feasibility. The semantic annotation facilitated the accurate service description, discovery of telecommunication network services and addressed the semantic interoperability problem. The proposed model-driven domain ontology modelling approach separates domain conceptual model from the concrete ontology modelling languages, it enhances the reusability of domain conceptual model and greatly reduces the technical difficulty of domain ontology modelling.

The remainder of this chapter is structured as follows. In Section 2 we presented a general domain ontology modelling methodology for telecommunications service field, and also proposed a specific model-driven domain ontology modelling approach to support the above presented methodology. Section 3 introduced the experimental environment and the demo service to validate the feasibility of domain ontology. Finally, conclusions are drawn.

2. Domain ontology modelling methodology for telecommunications service

Here, technical modelling details for the proposed approach are described, namely telecommunications service domain ontology modeling methodology and a corresponding model-driven implementation mechanism.

2.1 Domain ontology modeling methodology

Based on our practical experiences in recent years, a concrete domain ontology modeling methodology is summarized as shown in Figure 1. The modelling process is illustrated in detail as follows.

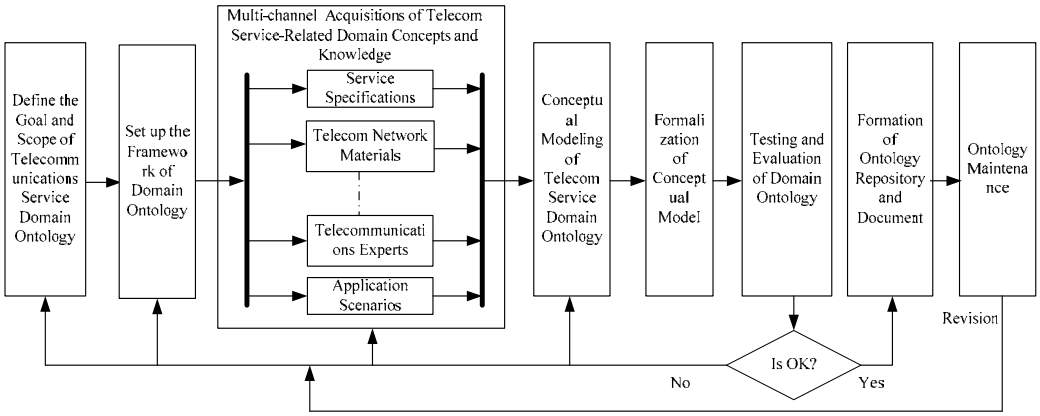


Fig. 1. Domain ontology modelling methodology.

2.1.1 Define the scope of telecommunications service domain ontology

The first step is mainly to define the scope and border of domain ontology. Telecommunications service domain ontology mainly addresses the semantic interoperability of telecommunications service. This domain ontology mainly provides the shared domain vocabularies and knowledge to support the semantic web applications in the telecommunication service field, such as semantic telecom service description, service discovery, and service context modelling. Therefore, TSDO should involve the service-related domain concepts and knowledge. For example, telecom services often involve network type, network carrier, billing policy, user terminal, service quality, service customer, service category, .etc. In fact, telecommunication service field consists of a large number of concepts/terminologies and relations. Some concepts have the higher sharing degree. However, some concepts are only related to concrete application field, such as service context ontology, service description ontology. Therefore, how to abstract the sharing domain concepts and reasonably organize them is a big challenge. The reusability and extensibility are two important ontology modeling factors considered. So an efficient ontology hierarchy modelling approach is needed.

In practice, we adopted a layered ontology modeling method to organize the domain concepts to improve the reusability and extensibility (see Figure 2). Common ontology, like time and space ontologies, can be shared in the different domains, like telecom, medical domain or any other domains. The concrete domain ontology can be shared by the different domain-related application ontologies. For example, TSDO may be used to create the service context ontology, network management ontology, etc. This method well distinguishes the border of TSDO, common ontology and telecom service-related application ontology.

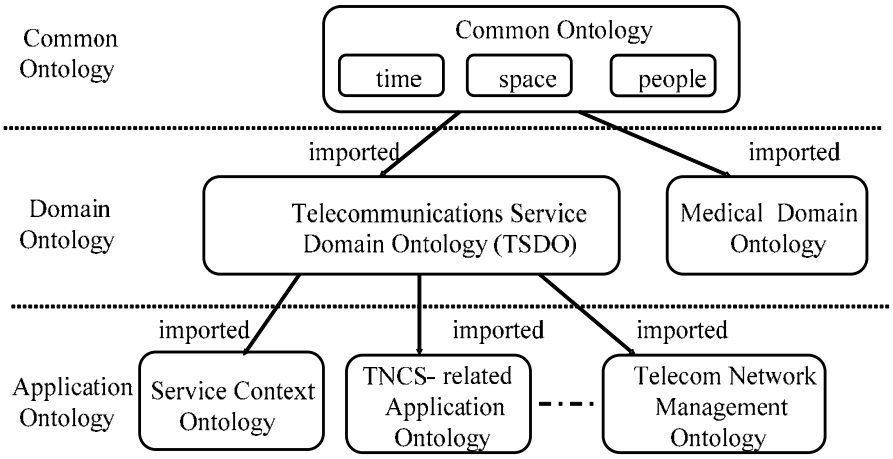


Fig. 2. Layered ontology modelling method.

2.1.2 Set the framework of telecommunications service domain ontology

When the goal and scope of TSDO are clear, the specific organization framework of TSDO should be set up. As TSDO involves a large number of telecom service domain concepts and relationships, how to reasonably classify and organize these terminologies is an important issue. Specifically, we adopted a modular modelling approach to construct TSDO. The principle of modular modelling is the "strong cohesion and loose coupling" way. The correlations among different concepts are the main reference of module division. The goal of modular modelling is to ensure that the correlation of concepts in the same module is stronger. Based on this modular design principle, TSDO is divided into several sub-ontologies as shown in Fig. 3.

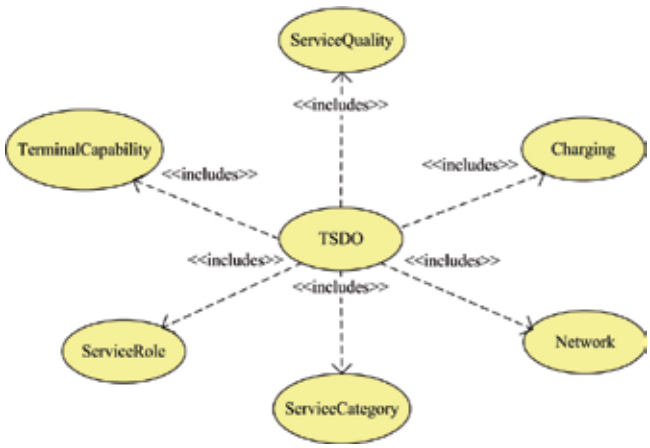


Fig. 3. The framework of telecommunications service domain ontology.

Specifically, TSDO mainly comprises six sub-ontologies, including Terminal Capability Ontology, Network Ontology, Service Role Ontology, Charging Ontology, Service Quality Ontology, and Service Category Ontology.

1. **Terminal Capability Ontology**: defines main concepts about terminal software, terminal hardware, terminal browser and network characteristics supported by terminal.
2. **Network Ontology**: specifies the network concepts, network category, network features, as well as the relationships of various networks, such as mobile network, internet, and fixed network, GSM, CDMA, UMTS, WCDMA, and WLAN.
3. **Service Role Ontology**: describes the stakeholders' concepts of the service supply chain, for example, service provider, content provider, network operator, service user.
4. **Service Category Ontology**: describes a telecommunications service classification. This ontology defines the relationship between various telecommunications services, like basic service, value-added service, voice service, data service, conference service, presence service, download service, browsing service, messaging service.
5. **Charging Ontology**: defines the charging-related concepts and rules about telecommunications services, including payment methods (such as prepaid and post-paid), charging types (such as time-based, volume-based, event-based, and content-based), billing rates, as well as account balances.
6. **Service Quality Ontology**: A telecommunication network must provide the services which have the end-to-end QoS guarantee. Depending on the technical characteristics, the QoS provided by different networks is varying. Service Quality Ontology mainly defines the QoS-related concepts about telecommunications service, including access network QoS, core network QoS and user's QoE, such as call delay, message size, call through rate, positioning accuracy, network bandwidth.

2.1.3 Multi-channel acquisitions of telecom service-related domain concepts and knowledge

After the framework of TSDO is set up, it needs to collect domain concepts and knowledge (including terminologies and their relationships) from multi-channel ways for each sub-ontology of TSDO. In general, the sources of knowledge acquisition include the released telecom service specifications, senior experts in the telecom field or some typical application

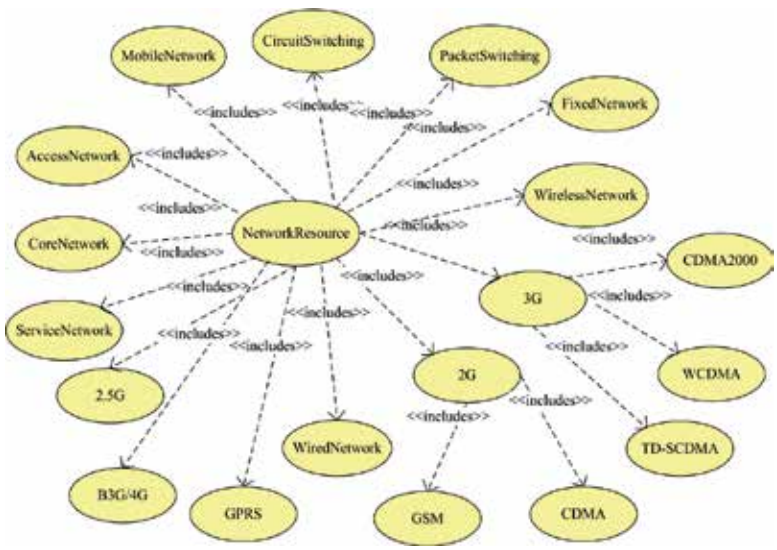


Fig. 4. Some collected domain concepts about telecom network.

scenarios. In this step, modellers need to list the collected concepts, relations and explanations as far as possible. It's unnecessary to care about the meaning overlap between the concepts and to consider how to express these concepts and their relation in class, property or instance ways. For example, Figure 4 briefly shows the concepts collection about network ontology.

2.1.4 Conceptual modelling of telecommunications service domain ontology

After the acquisition of a large number of telecom service related concepts, we need to make the concept classification, concept aggregation, and remove the duplicated concepts according to certain domain knowledge and logic. The goal of this step is to construct a conceptual model of TSDO. This concept model describes the involved domain concepts and their relations of each sub-ontology in detail. Note that, the relationships between the concepts not only involve the concepts of the same sub-ontology, may also be related to the concepts of different sub-ontologies. The concrete building of conceptual model is divided into three steps: (1) **Defining classes and class hierarchy**. In the process of defining the classes, we need to discover the inheritance hierarchy between the concepts and then distinguish the super-classes and sub-classes. (2) **Defining the properties of classes**. After the class is defined, its properties should be considered. There are two kinds of properties. One is datatype property, which is used to describe the features of the concept itself, such as name, age. The other is object property, which is used to depict the relationship between the concepts, like friendship relation between two people. (3) **The definition of domain axiom and knowledge**. When we use ontology to describe the real word things, there are often some contradictions or errors occurrences resulted by human negligence. For example, the range value of one person age property is negative, or a person has two biological fathers. To prevent these common-sense errors, some domain axiom and knowledge should be established. The axiom is to restrict the relationships of the concepts to ensure the consistency of domain knowledge, such as the range value or cardinality of properties.

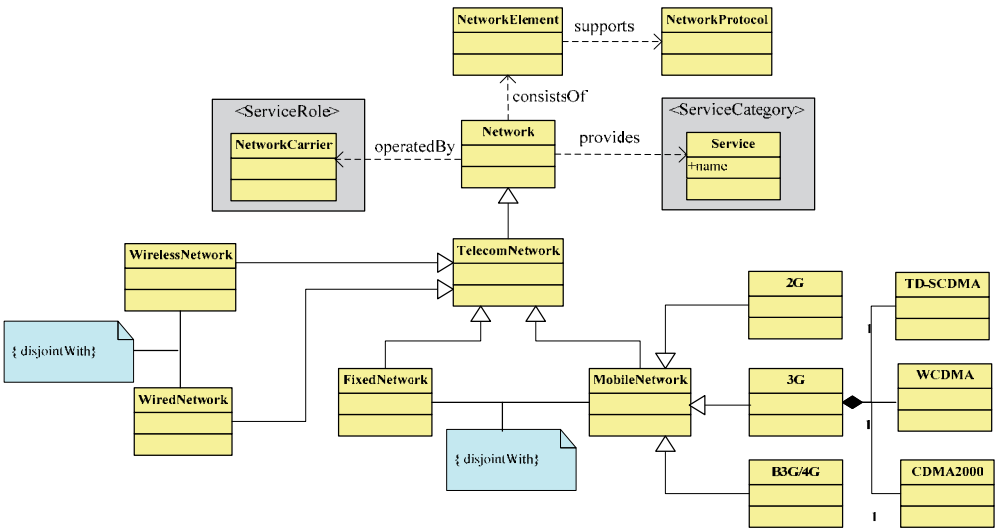


Fig. 5. Part conceptual model of network ontology.

Figure 5 shows the conceptual model of network ontology in part. Based on the terminologies collected in the above step, the class hierarchy and relationships are described. This conceptual model depicts the classification of network, the services provided by network and the operator of network. It can be seen that the ranges of object property "operatedBy" and "provides" are the concepts from ServiceRole and ServiceCategory sub-ontologies respectively. In addition, we define the domain axioms through the constraints way. For example, we define that "FixedNetwork" is disjointed with "MobileNetwork", i.e. if N1 is an instance of concept "FixedNetwork", then it will not be an instance of concept "MobileNetwork".

2.1.5 Formalization of conceptual model of telecommunications service domain ontology

As the conceptual model is one high-level abstract model and independent of any concrete ontology modelling languages, we need to formalize this conceptual model through a specific ontology modelling language like OWL (Web Ontology Language) (W3C, 2004a). In general, we can use the common ontology modelling tools to formally describe the terminologies, relationships and axioms in the conceptual model. Figure 6 shows the formalization description of part concepts and relationships of Figure 5 by OWL language. The concept is formally defined by "owl:Class", and the class hierarchy is organized by "owl:subClassOf". The "owl:ObjectProperty" is used to describe the relationships between the concepts and the "owl:disjointWith" clearly depicts the restrictions on the two disjointed concepts.

```
<owl:Class rdf:ID="Network"/>
<owl:Class rdf:about="#TelecomNetwork">
   <rdfs:subClassOf rdf:resource="#Network"/>
</owl:Class>
<owl:Class rdf:about="#MobileNetwork">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#TelecomNetwork"/>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#FixedNetwork"/>
</owl:Class>
<owl:Class rdf:ID="2G">
   <rdfs:subClassOf rdf:resource="#MobileNetwork"/>
  </owl:Class>
<owl:ObjectProperty rdf:ID="operatedBy">
   <rdfs:range rdf:resource="&ServiceRole;#NetworkCarrier"/>
   <rdfs:domain rdf:resource="#Network"/>
</owl:ObjectProperty>
```

Fig. 6. Part of network ontology formalized by OWL.

2.1.6 Evaluation of telecommunications service domain ontology

Ontology evaluation is an important issue that must be addressed if TSDO are to be widely adopted in the semantic related telecommunications applications. Ontology can be

evaluated against many criteria: its coverage of a particular domain and the richness, complexity and granularity of that coverage; the specific use cases, scenarios, requirements, applications, and data sources it was developed to address; and formal properties such as the consistency and completeness of the ontology. We can test and validate whether the domain ontology satisfy the requirement or not. If yes, these ontologies will be added to the ontology repository; if no, we have to return back to previous steps to make some revisions until the requirement is satisfied.

In the specific use process, we often can find some existing shortcomings of domain ontology. The utilization of domain ontology to formally describe the concrete application scenario is a very effective evaluation approach. For example, when we defined the TSDO, we use network, service role and service category sub-ontologies to describe the network carrier resource (see Figure 7). We found that the operating scope of network carrier is an important characteristic. But the concept "NetworkOperator" of service role sub-ontology lacks this property. Actually, some carriers can provide services through out nation; however, some carriers can only provide services in a specific province or region. Therefore, the property "CoverageScope" should be added to the concept "NetworkOperator" of service role sub-ontology.

```
<ServiceRole:NetworkOperator rdf:ID="ChinaMobileCommunicationOperator"/>
    <ServiceRole:CoverageScope rdf:resource="&LocationSpace;#TroughOutNation"/>
</ServiceRole:NetworkOperator>

<TelecomNetwork:GSM rdf:ID="ChinaMobileNetwork">
   <TelecomNetwork:operatedBy rdf:resource="#ChinaMobileCommunicationOperator"/>
   <TelecomNetwork:provides rdf:resource="&ServiceCategory;#DataService"/>
   <TelecomNetwork:provides rdf:resource="&ServiceCategory;#VoiceService"/>
</TelecomNetwork:GSM>
```

Fig. 7. Ontology description of china mobile communication operator.

2.1.7 Maintenance of telecommunications service domain ontology

The construction of domain ontology is the basis of ontology applications. However, as the different domain experts or ontology modelers may have the different understandings of the same domain concepts or relationships, some created ontologies may need to be further revised or improved in the practical utilization process. In addition, the knowledge of real world is growing and updated continuously. This also results that regular maintenance is necessary after ontoloies have been constructed. Ontology maintenance refers to a series of amendments, corrections, improvements and adaptive maintenance for ontology, which mainly consists of improving maintenance and adaptive maintenance. The improving maintenance is to revise or correct some existing errors of domain ontology. However, the adaptive maintenance refers to the extensions of existing domain ontology with the external real world changes, such as the knowledge increase or technology advances.

In addition, with the maturity of ontology technology, there are some ontologies developed by different research teams or communities to satisfy their different application needs. The main advantage of ontology is the knowledge sharing and reuse. How to realize the interoperation with these existent distributed ontologies is a big problem of ontology

maintenance. Therefore, sometimes, it needs to integrate several existent ontologies to address the reuse of different ontology knowledge. To implement the different ontology integration, the relationships among different ontologies should be analyzed. As the distributed feature and openness of WWW, knowledge ontologies maybe have the direct or indirect semantic relationships. For example, two ontologies maybe involve some same or similar concepts. The main relationships consist of two kinds: one is the repeat of terminologies definition. Some terminologies of this ontology might be equivalent to those defined in that ontology. It consists of the class equivalent and the property equivalent. For this equivalent relationship, we can use equivalent ontology mapping method to resolve as shown in Figure 8. The other is the subsumption of terminologies definition. It means that some terminologies of one ontology might subsume the semantic scope of those terminologies defined in other ontology. It also involves the class subsumption and property subsumption. For example, Figure 9 shows two independent ontologies: ontology 1 and ontology 2. In fact, the concept "Netowrk" of ontology 1 subsumes the concept "Internet" of ontology 2 in the semantic scope. Therefore, we can use the subsumption relationship to integrate these two ontologies into a new ontology.

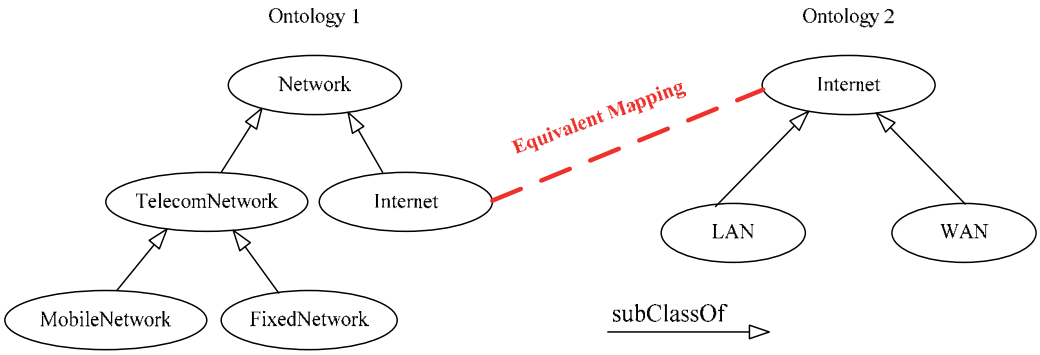


Fig. 8. Ontology integration based on the equivalent mapping.

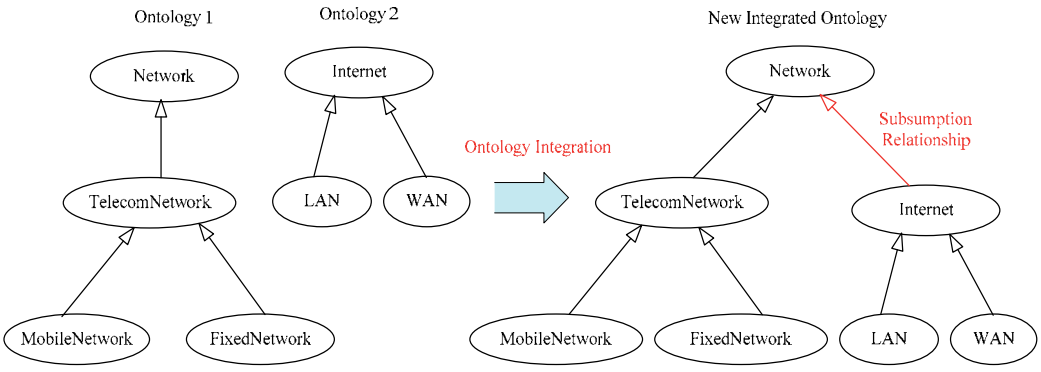


Fig. 9. Ontology integration based on the subsumption relationship.

2.2 A model-driven domain ontology modeling implementation approach

From the above descriptions in section 2.1, it can be seen that the construction of TSDO is a complex work, which involves not only several steps like terminology acquisition, concept modelling and formal description, but also different modellers like domain experts, formalization modeller. Currently, it lacks of a unified modelling tool to efficiently support this methodology. As the ontology modelling languages consists of a large number of logical symbols and formal description knowledge, it is not easy for general domain experts or software developers to understand and master. Although there are some visual modelling tools like Protege (Stanford, 2004) to support ontology modelling, the ontology modelling process still lacks the relation with mature software engineering method. For the general software developers, the current ontology modelling approach is not easy to master and it needs a strong professional background. Therefore, in the actual process of building domain ontology, domain experts often use UML (Unified Modelling Language) (OMG, 2005a) modelling tool or other office software to acquire domain terminologies or create concepts model, and then formalization modellers formalize the conceptual model by a specific ontology language through ontology modelling tool like Protege. As the existing UML modelling tool do not support the ontology modelling directly and the common ontology modelling tools also do not support the requirements and high-level conceptual modelling, the above proposed modelling process has to switch between different modelling tools. A key problem is that the high-level conceptual model cannot be automatically transformed into formal model encoded by a specific ontology language. This brings a lot of management and maintenance inconveniences of ontology modelling. The existing ontology modelling approach has limited the large-scale ontology development. Therefore, it needs a practical engineering approach and a unified modelling tool to support this modelling methodology completely.

Essentially, ontology engineering emphasizes the ontology modelling and knowledge reasoning; however, software engineering focuses on the complete system development methodology which mainly pays attention to requirement analysis, system design, implementation and dose not have the logical reasoning capability. So how to use mature software engineering theory and method to support the ontology development is very significant. Today, Model Driven Development (MDD) (Selic, 2003) is gaining significant momentum in both the software industry and the software engineering academic community. Model Driven Architecture (OMG, 2003), standardized by the Object Management Group (OMG), is a new strategy for designing software systems. Its main goal is to separate system function specification from specific implementation technique completely, enabling system's kernel function specification to be independent of the specific implementation platform technology. Therefore, MDA can retain the neutrality of programming languages, middleware platforms and vendors. In the face of heterogeneous and evolving technology, MDA is supposed to ensure: portability, increased application reuse and reduced development time. Thereby MDA minimizes the affection of technique changes.

Considering the development of domain ontology is a complex process and MDA is a new modeling approach which focuses on the model rather than the specific implementation technical details, we integrated MDA with ontology engineering together, and proposed a model driven domain ontology modeling approach to support the modelling methodology

described in section 2.1. By this approach, domain experts or general software developers, who are familiar with UML, can conveniently build the domain conceptual model by UML modelling tools and then this conceptual model can be automatically transformed into the corresponding ontology model encoded by a specific ontology language. As this approach separates domain conceptual model from the concrete ontology modelling languages like OWL, it enhances the reusability of domain conceptual model and reduces the technical difficulty of domain ontology modelling. The implementation details are described in the following sections.

2.2.1 Overview of model-driven TSDO modeling approach

MDA adopts the model-based development mode (Miller & Mukerji, 2003) as shown in Figure 10. Computation Independent Model (CIM) mainly describes the requirements of software system, which specify the system function and boundary. Platform Independent Model (PIM) is the high level abstraction of system function, without any information related to implementation techniques; Platform Specific Model (PSM) is the model which contains specific implementation platform technique information. The MDA–based development process is: firstly, establishing CIM based on the system requirements; secondly, according to the specifications of CIM, creating PIM with the platform independent modeling language, such as UML; thirdly, transforming the PIM to PSM according to some specific mapping rules; lastly, generating platform specific code automatically or semi automatically. In this process, modeller can further refine the created models in CIM, PIM or PSM stage.

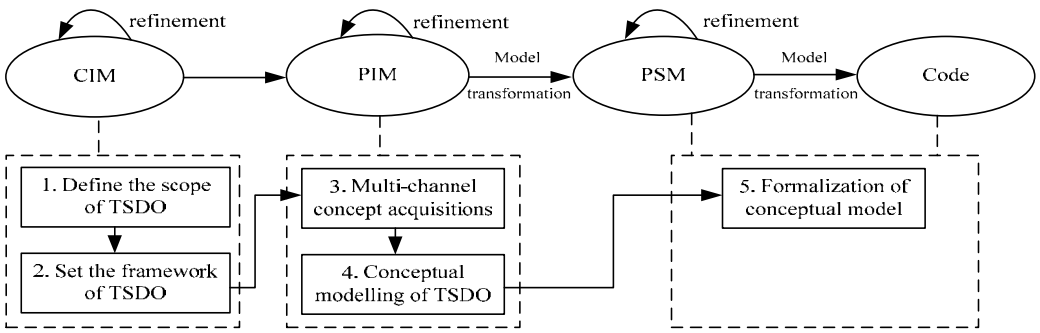


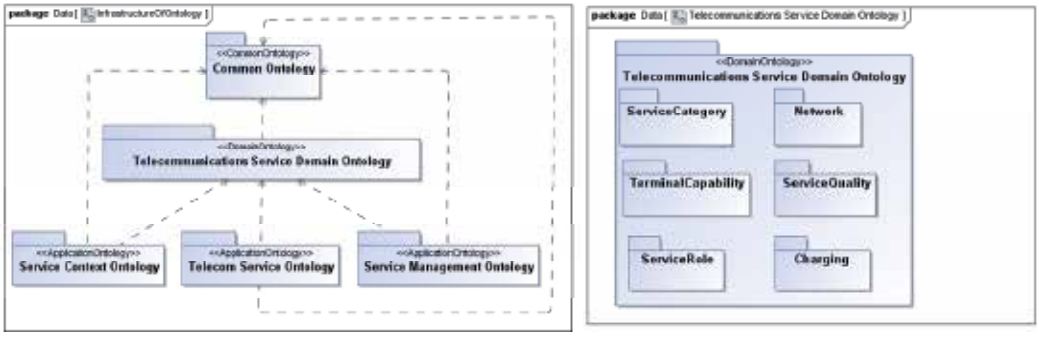
Fig. 10. Model-driven TSDO modelling approach.

According to the modelling idea of MDA, we presented a concrete model-driven TSDO modelling approach to provide a practical engineering implementation as shown in Figure 10. The definition of TSDO scope and the establishment of domain ontology framework belong to the CIM modelling stage. Modeller can employ use case diagram of UML to define the scope of TSDO and set up its framework. In this approach, PIM mainly focuses on the multi-channel domain concept acquisitions and the further conceptual integration and refinement, i.e. conceptual modelling. UML class diagram or use case diagram can be used to model the collected domain concepts and their relationships. After acquiring the domain terminologies, the following step is to integrate and refine these

concepts and their relationships to form a high-level domain ontology model, which is independent of a specific ontology description language. The PSM and code steps are used to realize the formalization of high-level domain ontology conceptual model by a specific ontology language. By the model to model transformation technology, the high-level conceptual model (i.e. PIM) can be transformed into an ontology language specific model (i.e. PSM). And then by using model to code transformation technology, the concrete ontology description file encoded by a specific ontology language like OWL (i.e. code) can be generated from the ontology language specific model (i.e. PSM). When we need to revise or maintain the created ontology, we can return back to the CIM or PIM to modify the related models and then generated the corresponding code again. In this mode-driven ontology development approach, all processes adopt the standard UML model or UML extension mechanism (i.e. UML Profile). The technical details are described in the following sections.

2.2.2 CIM step: The scope and framework modeling of TSDO

In order to well organize the development of TSDO, this approach uses the UML use case diagram to model the scope and framework of TSDO. As is shown in Figure 11(a), the ontology hierarchy is represented by package *InfrastructureOfOntology*, which consists of three types of package: common ontology, domain ontology and application ontology. Each package contains the related ontology concepts and their relationships. The *Common Ontology* package contains some general concepts particularly designed for high reusability, where other different domain ontologies and application ontologies either import or specialize its specified concepts or relationships. This is illustrated in Figure 11(a), where it is shown how domain ontologies and application ontologies each depends on the common ontology. The common ontology is generally defined by some standard organizations or research communities. In this chapter, we mainly focus on the building of telecommunications service domain ontology. In order to facilitate reuse, the *Telecommunications Service Domain Ontology* package is further subdivided into a number of packages: *ServiceCategory, Netowrk, TerminalCapability, ServiceQuality, ServiceRole*, and *Charging*, as shown in Figure 11(b).

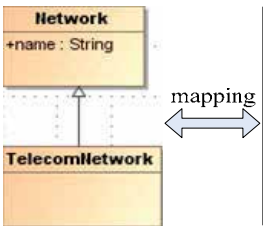


(a) (b)

Fig. 11. CIM modelling of telecommunications service domain ontology.

2.2.3 PIM step: Terminology acquisitions and conceptual modeling of TSDO

After defining the CIM of TSDO, the following step is to construct the PIM of TSDO. It means that the telecom service related domain terminologies should be collected and then integrated into a high-level abstract domain ontology model which is independent of a specific ontology language like OWL. The collection of domain terminologies can be modeled by the UML Use Case diagram like Figure 4. However, the high-level domain conceptual modeling is the emphasis of PIM. How to model the conceptual model of domain ontology based on UML is needed to resolve. Fortunately, UML and ontology language have some common features, although sometimes represented differently. This provides a possible transformation from UML model to ontology model. For example, both ontology representation language and UML are based on *Class*. The *Generalization* elements of UML can represent the subClass or subProperty semantic of ontology. The ownedAttribute of UML Class can describe the DatatypeProperty of ontology language. The mapping example is illustrated in the Figure 12.



mapping

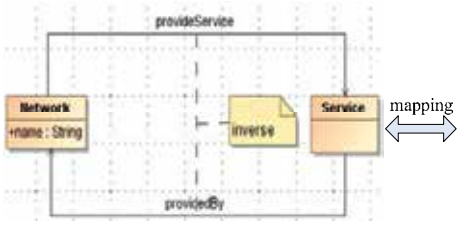
```
<owl:Class rdf:ID="Network"/>
 <owl:Class rdf:ID="TelecomNetwok">
  <rdfs:subClassOf rdf:resource="#Network"/>
 </owl:Class>
 <owl:DatatypeProperty rdf:ID="name">
  <rdfs:domain rdf:resource="#Network"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
 </owl:DatatypeProperty>
```

(a) (b)

Fig. 12. The direct mapping example from UML to OWL.

However, although UML Class diagram has some constructs similar to the constructs of ontology representation language, there are still some ontology constructs which cannot be represented by UML constructs directly. We need to find the appropriate UML elements to represent some other ontology constructs, like objectProperty, equivalent class relation, and disjointing class relation. For instance, we can select the directedAssociation element of UML to represent the ObjectProperty and use the constraints anchored with association to represent the inverse, symmetric or transitive feature of ObjectProperty. An illustrated example is shown in Figure 13.

As a common software modeling language, most of software developers, system analysts and designers are familiar with UML. So, in order to decrease the technical threshold, it's a practical approach for the conceptual modeling of TSDO by UML. Although UML has some similar constructs with ontology language, however, the modeling goals and description capabilities of both languages have some differences. From the above analysis, in order to use UML to represent high-level ontology conceptual model, we need to define a specific tailored representation method to guide the modeler to build the conceptual model of domain ontology. Table 1 shows the main corresponding relation of UML elements with ontology elements. According to this semantic representation way, the modeler can use the UML elements to describe the semantic-enabled high-level ontology conceptual model like Figure 14.



mapping

```
<owl:Class rdf:ID="Network"/>
<owl:Class rdf:ID="Service"/>
<owl:ObjectProperty rdf:ID="providedBy">
  <rdfs:range rdf:resource="#Network"/>
  <rdfs:domain rdf:resource="#Service"/>
  <owl:inverseOf>
    <owl:ObjectProperty rdf:ID="provideService"/>
  </owl:inverseOf>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#provideService">
  <rdfs:domain rdf:resource="#Network"/>
  <rdfs:range rdf:resource="#Service"/>
  <owl:inverseOf rdf:resource="#providedBy"/>
</owl:ObjectProperty>
<owl:DatatypeProperty rdf:ID="name">
  <rdfs:domain rdf:resource="#Network"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

(a) (b)

Fig. 13. The indirect mapping example from UML to OWL.

UML Elements	Ontology Elements	Comments
Class	Class	
Generalization	subClass, subProperty	
Instance	Individual	
Multiplicity	minCardinality maxCardinality	ontology cardinality declared only for range
ownedAttribute	Datatype Property	
directedAssociation	ObjectProperty	The value of "owned By" property of Association End A is the domain of ObjectProperty, the value of "owned By" property of Association End B is the range of ObjectProperty.
Constraint	Inverse Symmetric Transitive Functional	
Enumeration	oneOf	
Association Class	disjointWith equivalentClass	

Table 1. The defined UML representation method for high-level conceptual model of domain ontology.

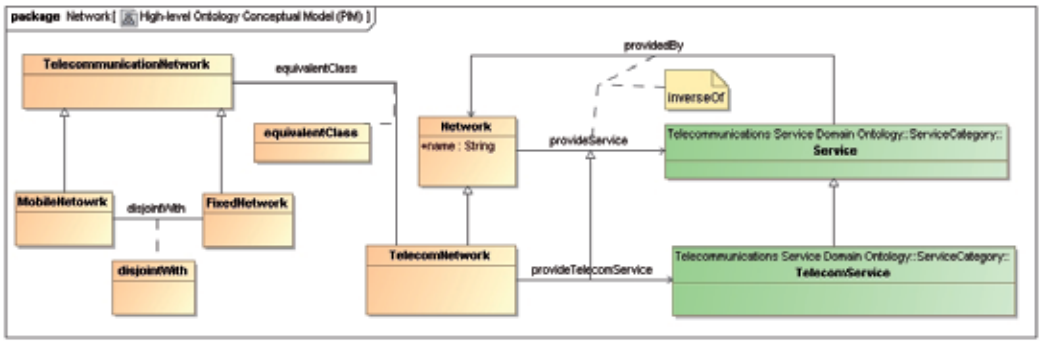


Fig. 14. PIM: A part of high-level conceptual model of network ontology.

2.2.4 PIM to PSM step: Formalization of ontology conceptual model

It can be seen that the high-level ontology conceptual model described by UML is independent of a specific ontology language. So, in order to generate the formal file encoded by a specific ontology language, we need to transform the PIM into PSM according to the concrete model transformation rules. Figure 15 shows the general model transformation mechanism of model driven architecture. Model transformation is essentially to map the source model elements to other elements of the target model. Models are usually the instantiation of its meta-model. The model transformation rules are generally defined in the metamodel level and then model transformation engine apply these rules to the model level to complete the model transformation.

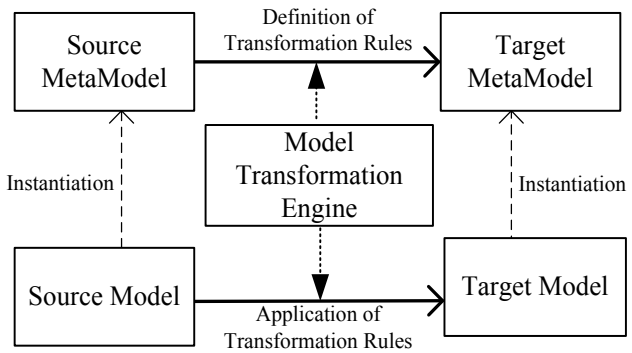


Fig. 15. The principle of model transformation.

Therefore, in order to transform the high-level ontology conceptual model (i.e. PIM) into platform specific model (i.e. PSM), we need to define the transformation rules according to the source and target metamodels. In our proposed approach, the high-level ontology conceptual model (i.e. PIM) is modeled by UML2.0, and the source metamodel is UML2.0 metamodel obviously. So we need a target metamodel relating to specific ontology language to describe the formal ontology model (i.e. PSM). In fact, OMG (Object Management Organization), which is the promoter of MDA, has considered this problem. In May 2009, OMG released the Ontology Definition Metamodel (ODM) v1.0 (OMG, 2009) based on the

meta-modeling mechanism of MDA. This specification represents the foundation for an extremely important set of enabling capabilities for MDA based software engineering, namely the formal grounding for representation, management, interoperability, and application of business semantics. The ODM is applicable to knowledge representation, conceptual modeling, formal taxonomy development and ontology definition, and enables the use of a variety of enterprise models as starting points for ontology development. ODM is based on the Meta Object Facility (MOF) (OMG, 2006) meta-modeling architecture of MDA, illustrated by Fig.16, which is based on the traditional four layer metadata architecture. From top to bottom, meta-data is abstracted to 4 layers: M3 (meta-meta model), M2 (meta model), M1 (model) and M0 (object and instance). The under-layer is the instance of its up-layer in turn. M3 layer is the end of meta-layer, namely, MOF is self-described. MOF is a common, abstract language used to define meta-model. It defines some meta-modeling constructs, such as Class, DataType, Association, Package, and Constraint. So the meta-model of ODM or UML can be defined by MOF, whose power just lies in its capability to enable interoperability among different meta-models. Currently, there are 2 approaches to construct meta-models in M2 layer. One is to make use of MOF to define a completely new meta-model from syntax to semantics. Although this approach supports to define a new meta-model that will perfectly match the concepts and relation of the concrete domain, this need the underlying programming realization of corresponding new modeling tool. This is heavy-weight meta-modeling, such as UML and ODM. The other is to extend the existent UML meta-model and then construct a standard UML Profile through UML extension mechanism (Stereotype, TaggedValue, Constraints). This approach allows both defining domain specific conception and relation through UML extension mechanism and using the intrinsic UML elements. So it's a light-weight meta-modeling approach and most of existent MDA tools support this UML Profiling-based meta-modeling mechanism currently. There is no need to develop a new modeling tool. From the above analysis, the UML Profiling-based meta-modeling mechanism approach is adopted in our approach.

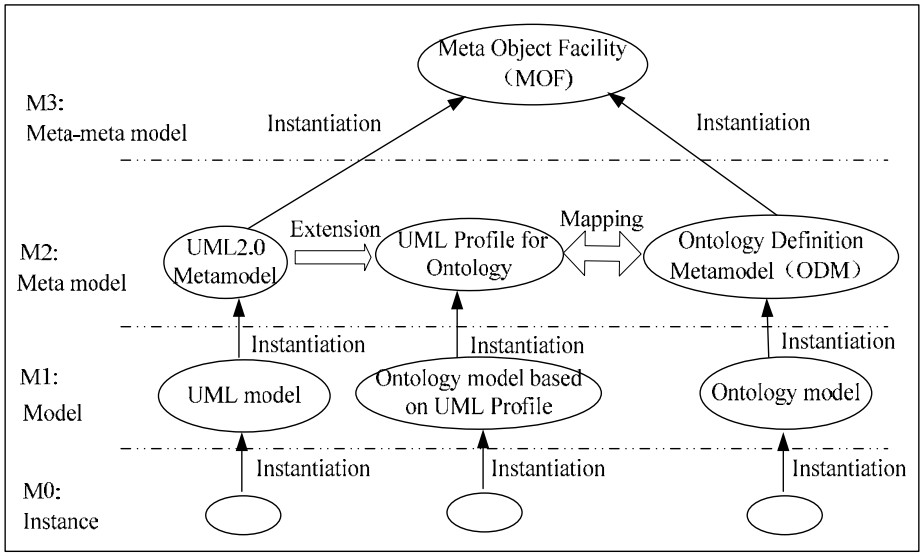


Fig. 16. ODM: the integration of semantic web and model driven architecture.

Therefore, in this approach, the metamodel of PSM employs the UML Profile for RDF and OWL defined in ODM specification. This profile is designed to support modelers developing vocabularies in Resource Description Framework (RDF) (W3C, 2004b) and richer ontologies in the Web Ontology Language (OWL) through reuse of UML notation using tools that support UML2 extension mechanisms. Table 2 specifies a part of stereotypes set that comprise the UML2 Profile for using UML to represent RDF/S and OWL vocabularies.

RDF、RDFS and OWL ontology	**UML Base Class**	**UML Stereotype**
rdfs:Resource	Class	《rdfsResource》
rdfs:Datatype	Class	《rdfsDatatype》
rdfs:domain	Association	《rdfsDomain》
rdfs:range	Association	《rdfsRange》
rdfs:subClassOf	Generalization	《rdfsSubClassOf》
rdfs:subPropertyOf	Generalization	《rdfsSubPropertyOf》
owl:Class	Class	《owlClass》
owl:Restriction	Class	《owlRestriction》
owl:ObjectPropert	Class AssociationClass Property Association	《objectProperty》
owl:DatatypeProperty	Class AssociationClass Property Association	《datatypeProperty》
owl:equivalentClass	Constraint	《equivalentClass》
owl:disjointWith	Constraint	《disjointWith》

Table 2. A part of UML Profile for RDF and OWL.

After the source and target metamodels are determined, we can define the model transformation rules from high-level ontology conceptual model (i.e. PIM) to ontology language related model (i.e. PSM). For example, based on the Table 1 and Table 2, we can define the following model transformation rules to support the model transformation like Figure 17. Notably, the source metamodel is UML2.0 metamodel and the target metamodel is UML Profile for RDF and OWL in this proposed approach.

When the transformation rules are defined, the model transformation engine can scan the elements of source model and then transform them into the corresponding elements of target model according to the transformation rules. As model transformation is a key technique used in model-driven architecture. In 2002, OMG issued a Request for proposal (RFP) on MOF Query/View/Transformation to seek a standard compatible with the MDA recommendation suite (UML, MOF, OCL, etc.). Several replies were given by a number of companies and research institutions that evolved during three years to produce a common proposal that was submitted and approved. QVT (Query/View/Transformation) (OMG,

2008) is a standard set of languages for model transformation defined by the Object Management Group. Currently, some MDA tools have declared to support the complete or part functions of QVT. For example, by using the transformation rules, the source model in Figure 14 is transformed into a target model in Figure 18.

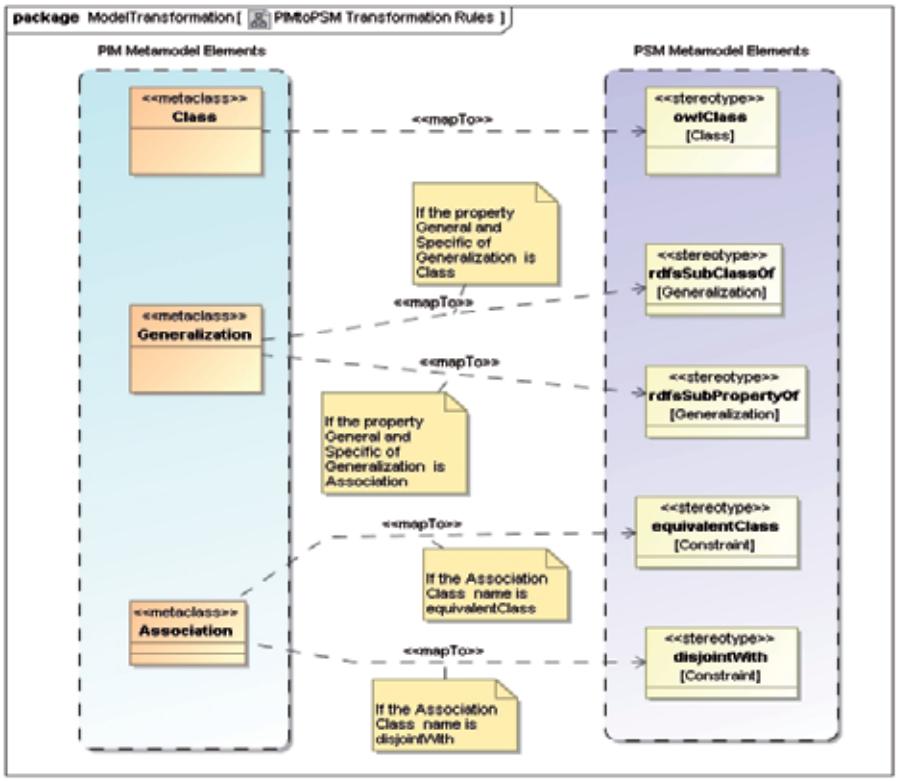


Fig. 17. A part of PIM to PSM transformation rules definitions.

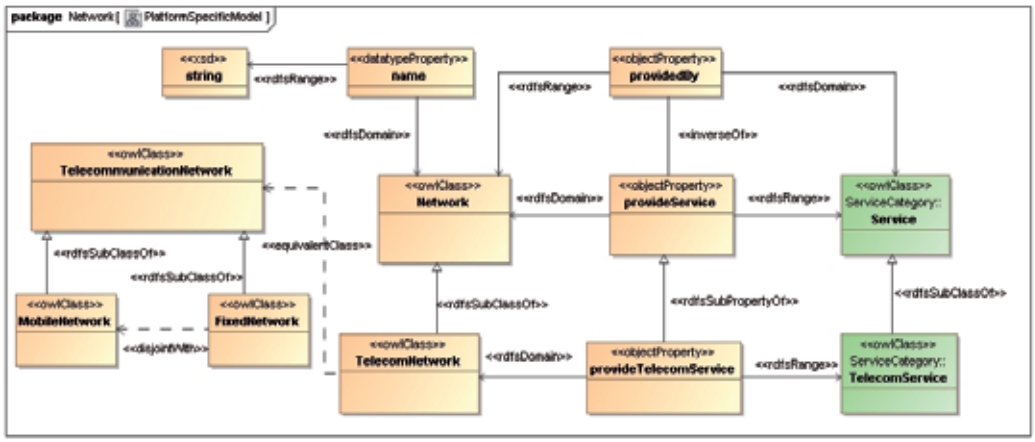


Fig. 18. PSM: A part of ontology specific model based on UML Profile for RDF and OWL.

2.2.5 PSM to code step: Formalization of ontology conceptual model

In order to generate the formal ontology file encoded by OWL, the PSM based on UML Profile for RDFS and OWL should be transformed into ontology file encoded by OWL, which is a model to code transformation process, which involves the model scanning technology.

2.2.5.1 Model to code transformation theory

Before we introduce the concrete transformation process, some related definitions are given firstly.

Definition 1. *Ontology model triples: UmlOnt (C, R, G).*

PSM based on the UML class diagram can be represented by a triple: UmlOnt(C, R, G). C is the class node set, and it is an ontology class definition of the concept in PSM. R is the relation node set, and it is the definition of the relations among ontology class. G is the relation set of C and R, which describes the relations among the nodes in C and R set.

Definition 2*. Relation Matrix (RM)*

Relation matrix is a n order square matrix, including elements a_{ij} in total of n * n, which looks like $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$, denoted as $A = (a_{ij})_{nn}$ (i,j = 1, 2, 3, ……, n) is used to describe the relations between the class nodes and relation nodes in the PSM. And a_{ij} indicates whether there is relation between class node *i* and class node *j,* as well as the type of relation node.

Definition 3*. Connected subgraph*

Given a directed graph, which can be divided to several connected subgraphs {G1, G2, ⋯, Gn}, these subgraphs meet the following conditions:

- In any two connected subgraphs Gi and Gj, there is no such a node x which is both in Gi and Gj at the same time. That is, $\neg\exists x, x \in Gi \cap x \in Gj$.
- In a connected subgraph Gi, there always exist directed edge between any two nodes x and y (without regard to the direction of the edge).

Definition 4*. Model Transformation Automaton (MTA)*

Model transformation automaton is a quintuple: MTA = (Q, Σ, δ, q_0, F), including:

- Q: A nonempty and finite set of states and one state in it corresponds to an ontology class node. ∀ q∈Q, q is called a state of MTA.
- Σ: Input events table, in which one input event corresponds to an ontology relation node.
- δ: Transfer function. One transfer function corresponds to a nonzero number a_{ij} ($a_{ij} \neq 0$), δ：Q×Σ→Q in relation matrix.
- q_0: The begin state of MTA, q_0∈Q.

- F: The set of terminate states. F is included by Q. Any q∈F, q is called a terminate state of MTA.

According to the definitions given above, the transformation engine from PSM to formal file encoded by OWL can be described as: The model transformation engine firstly scans the PSM class graph, and the scanning result generates the ontology model triples UmlOnt(C, R, G). C is the set of all nodes in UML class graph, R is the set of all relations in UML class graph, G represents the structure relationship of the ontology class graph, which can be regarded as N connected subgraphs divided from a directed graph and these subgraphs correspond to N relation matrices {RM1, RM2, ⋯, RMn }. One nonzero number a_{ij} represents the relation type between the class node i and j, and these relations are all included in R.

When the transformation engine finishes scanning, it input the scan result to the model transformation automaton. In this MTA, the nonempty finite set of states corresponds to C in the ontology model triples; the input events table corresponds to R in the ontology model triples; the transfer function corresponds to G in the ontology model triples; q0 and F are elements in C. In the procedure of state transforming, the corresponding operations of model transformation are also performed in MTA. When the automaton arrives at the terminal, the transformation finishes.

2.2.5.2 The implementation mechanism of model to code transformation engine

In order to realize the model to code transformation according to the above mentioned theory, we design a model transformation engine based on Eclipse Plugin technology. In MDA, XML Metadata Interchange (XMI) (OMG, 2005b) is an Object Management Group (OMG) standard for exchanging metadata information via Extensible Markup Language (XML). As the most of MDA tools use XMI as an interchange format for UML models, the model transformation engine is responsible for scanning PSM encoded by XMI and then transforming PSM into ontology file encoded by OWL. The process of model to code transformation is indicated in Figure 19.

2.2.5.2.1 Model scanning module

When building ontology model, different modeling tool means different element label and different label structure in the model description file. Therefore, this chapter proposed a transitional model convert method, which adopts same data structure when describes different model format, i.e. the triples in *Definition 1*. This allows the model transformation is no longer constrained by the model structure. It thereby improves the versatility of transformation engine and is convenient to be maintained and updated.

The model scanning module in transformation engine scans the UML class graph encoded by XMI, and the scan result will generate two list sets in the transitional model. It is used to store the class nodes and relation nodes of UML graph, which corresponds to the C and R set in the UML ontology model triples.

2.2.5.2.2 Building relation matrix module

The function of relation matrix building module is used to generate the relation matrix of PSM, i.e. the G set in UML ontology model triples. There are mainly two kinds of nodes in

UML class graph: class node and relation node. Relation node connects class node and distinguishes them by direction, which is very similar to the directed graph. Hence directed graph is adopted to represent UML class graph. Usually matrix is used to represent the graph, and the values in matrix represent the type of relation. In ontology model class graph, there may be several independent subgraphs, which satisfy the description in definition 2. Therefore, it is necessary to handle the relation matrix to produce N independent sub relation matrices. This method can reduce the order or relation matrix and thereby reduce store space of the model, which also improves the efficiency of model transformation.

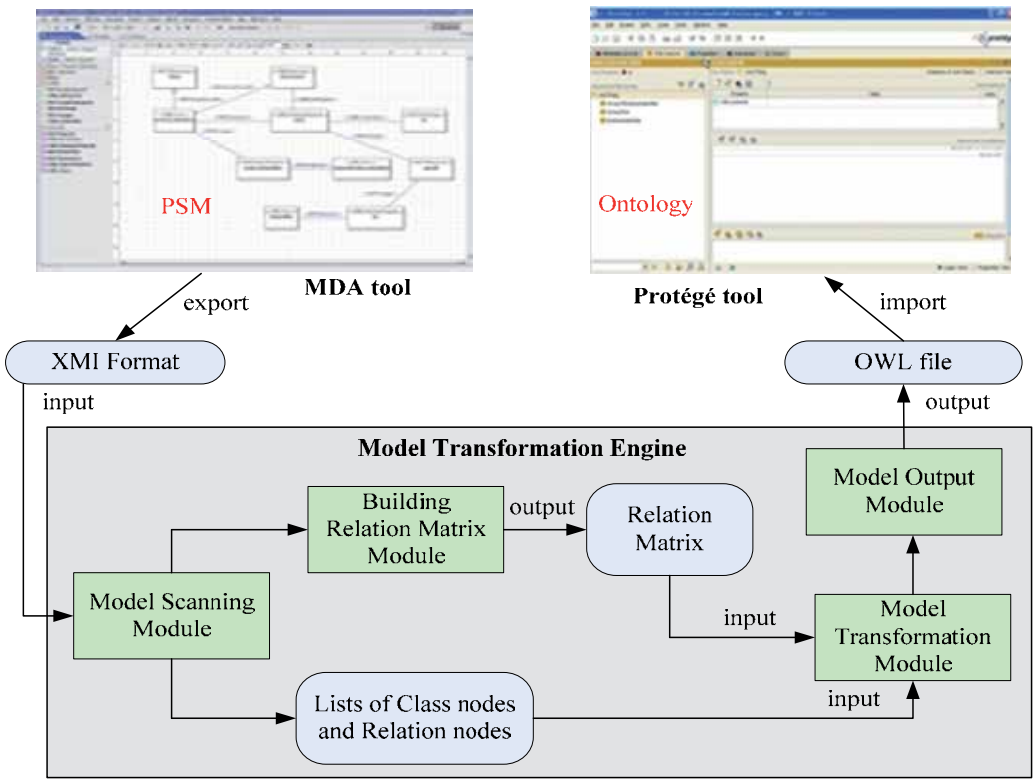


Fig. 19. Model to code transformation process.

2.2.5.2.3 Model transformation module

Model transformation module includes model transformation automaton and model transformation regulation table. In the process of states transition, the automaton performs transformation from PSM to OWL according to the corresponding transformation regulations. In this module, the automaton is separated with the model transformation regulations. Therefore, the changes of model transform regulation will not influence the running of automaton, and it is convenient to perform daily maintenance and update of the engine.

The model transformation automaton is a quintuple, and all information of this quintuple are included in the model transform transitional model, namely in the UML ontology

description model triples (C, R, G). The states, input events and transformation function of the automaton correspond to the class nodes, relation nodes and relation matrix set respectively in PSM class graph. The begin state of automaton is the owlClass node or objectProperty/datatypeProperty node in class nodes and the terminate states set includes all class nodes whose out-degree are zero and all nodes have been transformed by the automaton.

The model transformation regulation table defines the transformation regulation from UML Profile for RDF and OWL to OWL language. In the states jump process of model transformation automaton, corresponding regulation is used to perform model transformation. In the process of formulating the transformation regulation, the relations between every label node should be unified, which makes the regulation can be formulated depending on the OWL label structure and the relations between UML model elements. And good regulation is easy to extend in the future.

2.2.5.2.4 Model output module

Model output module only stores the formalized result of the transformation to the appointed path. And in order to verify the validity of the OWL file transformed, user can import the generate code into protégé tool for verification. Protégé is an ontology editor developed by Stanford University, which represents the OWL structure in graphic interface and makes the verification of OWL code validity more quickly and conveniently.

By using the above mentioned model to code transformation approach, the PSM of Figure 18 is transformed into the corresponding ontology encoded by OWL like Figure 20.

3. Experimental environment, use cases and evaluation

In this section, we describe our experimental environment, the implemented service use case and present the obtained evaluation results to validate the semantic interoperability enabled by telecommunications service domain ontology.

3.1 Experimental environment

In order to support this model-driven domain ontology modelling approach, Borland Together (Borland, 2006), a famous MDA tool, is employed in our experiment. By using UML extension mechanism, we implemented the UML Profile for RDF and OWL in Borland Together. In addition, Borland Together tool enable the model-to-model transformation, and this facilitates the transformation from PIM to PSM. Through the developed model-to-code transformation engine, we realize the transformation from PSM to ontology file encoded by OWL. At last, in order to verify whether the transformation is correct or not, the generated OWL file is imported in Protégé tool to test. By the experimental verification, the proposed model-driven ontology modelling approach can nicely support the constructing methodology of telecommunications service domain ontology.

Under the guidance of this approach, our research team has created a telecommunications service domain ontology knowledge repository which consists of around 430 telecommunications services-related ontology concepts/terminologies and 245 properties. Currently, these ontologies are published on our website (BUPT, 2009), see Figure 21.

```
<owl:Class rdf:ID="MobileNetwork">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="TelecommunicationNetwork"/>
  </rdfs:subClassOf>
  <owl:disjointWith>
    <owl:Class rdf:ID="FixedNetwork"/>
  </owl:disjointWith>
</owl:Class>
<owl:Class rdf:ID="TelecomNetwork">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="Network"/>
  </rdfs:subClassOf>
  <owl:equivalentClass>
    <owl:Class rdf:about="#TelecommunicationNetwork"/>
  </owl:equivalentClass>
</owl:Class>
<owl:Class rdf:about="#FixedNetwork">
  <owl:disjointWith rdf:resource="#MobileNetwork"/>
  <rdfs:subClassOf>
    <owl:Class rdf:about="#TelecommunicationNetwork"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#TelecommunicationNetwork">
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
  <owl:equivalentClass rdf:resource="#TelecomNetwork"/>
</owl:Class>
<owl:ObjectProperty rdf:ID="provideTelecomService">
  <rdfs:range rdf:resource="&ServiceCategeory; #TelecomService"/>
  <rdfs:domain rdf:resource="#TelecomNetwork"/>
  <rdfs:subPropertyOf>
    <owl:ObjectProperty rdf:ID="provideService"/>
  </rdfs:subPropertyOf>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="prvidedBy">
  <rdfs:domain rdf:resource="&ServiceCategeory; #Service"/>
  <rdfs:range rdf:resource="#Network"/>
  <owl:inverseOf>
    <owl:ObjectProperty rdf:about="#provideService"/>
  </owl:inverseOf>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#provideService">
  <rdfs:range rdf:resource="&ServiceCategeory; #Service"/>
  <owl:inverseOf rdf:resource="#prvidedBy"/>
  <rdfs:domain rdf:resource="#Network"/>
</owl:ObjectProperty>
<owl:DatatypeProperty rdf:ID="name">
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  <rdfs:domain rdf:resource="#Network"/>
</owl:DatatypeProperty>
```

Fig. 20. A part of formal network ontology encoded by OWL.

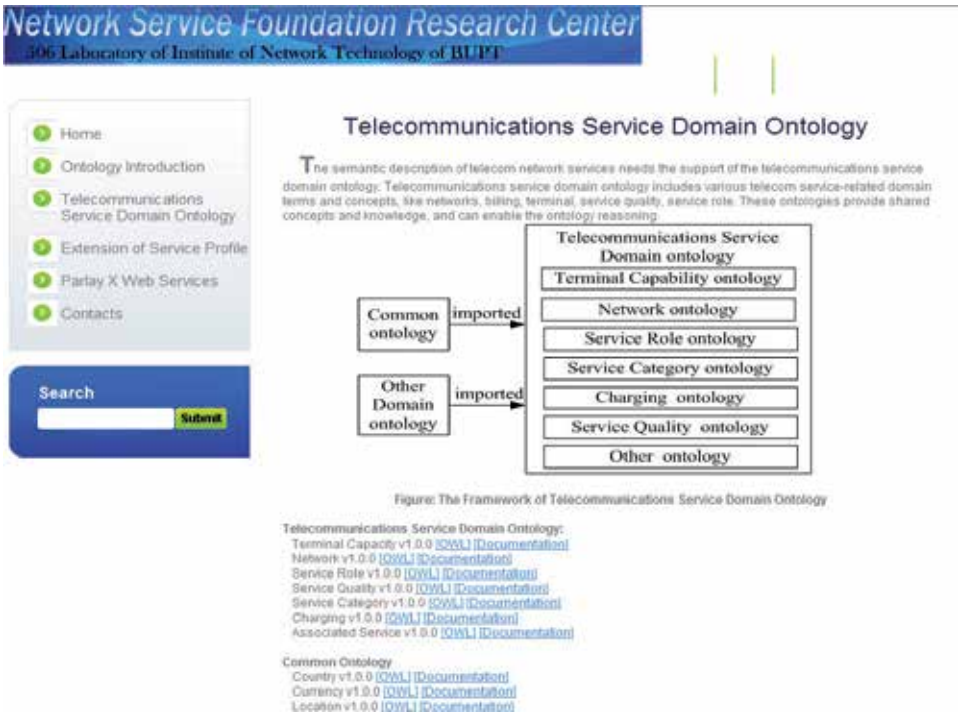


Fig. 21. The published telecommunications service domain ontology.

3.2 Use cases: Semantic telecommunications network capability services

In order to support the shift from traditional closed business model to open service ecosystem of telecom industry, NGN (Next Generation Network) and 3G network all adopt the open API (Application Programming Interface) technologies in the service layer, such as Parlay/OSA and Parlay X (Moerdijk & Klostermann, 2003). Thus, the telecommunication network services, such as call control, short messaging service, and location service, are available to the service developers in the form of APIs. This facilitates the value-added service development. With the development of distributed computing technology, Service-Oriented Architecture (SOA) is also imported into the telecommunications service domain by Parlay Web Service specifications. However, the open interface specifications of telecommunication networks are currently still in the syntactic level. As WSDL (Web Services Description Language)-based telecommunication network services lack the rich semantic annotation information, the keyword-based service matching cannot enable an accurate service discovery. So, currently value-added services often directly invoke the needed telecom network services provided by a specific network carrier. This results in the tight-coupling of application logic and service resources, which limits the provision of dynamically self-adaptive services. The applications cannot dynamically discover satisfied telecom network services and compose them according to the context environment. Facing the heterogeneous networks and personalized user demands, the self-adaptation has become a very important feature of future intelligent integrated service. Therefore, the semantic interoperability of telecom network and Internet in the service layer should be considered.

Based on this domain ontology, we described the telecom network capability services in the semantic level to validate its feasibility. We apply the semantic web service and ontology technologies to the telecommunications service domain, and present an infrastructure to enable the semantic interoperability of telecom network and Internet in the service layer (Qiao et al., 2008b). The proposed approach improves the accuracy of telecommunication network services description, discovery and matching, and unifies the semantic representations of telecommunication and Internet services.

3.3 Lessons learned

Currently, under the shift trend from Web2.0 to Web3.0 era, there have been some initial semantic web applications in Internet field. For example, the system of Twitter allows tweets to be tagged with information that will not appear in the message but can be read by computers (Twitter, 2010). Google is using structured data open standards such as microformats and RDFa to power the rich snippets feature. It's an experimental Semantic Web feature (Google, 2010). FOAF (Friend of a Friend) (FOAF, 2010) is a machine-readable ontology describing persons, their activities and their relations to other people and objects. As a "practical experiment" in the application of RDF and Semantic Web technologies to social networking, FOAF is becoming more and more popular now (FOAF, 2000). In addition, Linked Data (Linked Data, 2007) is a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF.

However, the semantic web applications in telecommunication services domain are still in an early research phase. Although RDF-based CC/PP (Composite Capability/Preference Profiles) (W3C, 2007) and UAProf (User Agent Profile) (OMA, 2001) are used to describe the terminal capability and user preference, other practical applications are very rare. Therefore, in order to eliminate the semantic gap between telecom network and Internet, the research on semantic web applications in telecommunications field still need to be further enhanced. Telecommunications service domain ontologies consist of various domain related concepts and knowledge, which is the base of semantic interoperability. The wide acceptance of standards and common practices of telecommunications service domain ontologies are still a way ahead. The promotion of the telecommunications service domain ontology by related standardization organizations would be in the foundation for the semantic interoperability of heterogeneous communications equipments and the industrial practical convergent service integration.

4. Conclusion

The network heterogeneity and service convergence are the main characteristics of future network. The provision of self-adaptive intelligent integrated services has become the pursuing goal of network carriers and value-added service providers. Dynamic discovery and composition of services are the important enabling technologies for self-adaptive integrated services. In the service discovery and composition process, semantic interoperability is a key issue. Actually, ontology, as a semantic interoperability and knowledge sharing foundation, has obtained more and more attentions. However, telecommunication service field consists of a large number of concepts/terminologies and

relations. How to abstract the sharing domain concepts and reasonably organize them is a big challenge. In this chapter, we presented a practical domain ontology modelling approach for telecommunications service field. Based on this approach, we constructed an open telecommunication service domain ontology repository to support the knowledge sharing and reuse. This will partly facilitate the semantic interoperability of the telecommunications networks and the Internet in the service layer.

5. Acknowledgment

This work was supported by National Key Basic Research Program of China (973 Program) under Grant No. 2012CB315802, National Natural Science Foundation of China under Grant No. 60802034, No. 61171102 and No. 61132001, Beijing Nova Program under Grant No. 2008B50 and New generation broadband wireless mobile communication network Key Projects for Science and Technology Development under Grant No. 2011ZX03002-002-01. We thank Huawei Technologies Co., Ltd. for cooperation in promotion of this work. Thanks also to Dr. Anna Fensel, a Senior Researcher at FTW – Telecommunications Research Center Vienna and STI Innsbruck, University of Innsbruck, Austria, for her valuable comments and suggestions.

6. References

Bashah, N.S.K., Jorstad, I. & Thanh, D.V. (2010). Service Discovery in Future Open Mobile Environments. *Proceedings of ICDS 2010 Fourth International Conference on Digital Society,* pp.47-53, ISBN 978-1-4244-5805-9, St. Maarten, Netherlands Antilles, February 10-16, 2010

Borland, (2006). Together: Visual Modeling for Software Architecture Design. 26.09.2011, Available from http://www.borland.com/us/products/together/

BUPT, (2009). Semantic web application for telecommunications service. 26.09.2011, Available from http://www.int.bupt.cn/jsp/centers/bupt506/intro.htm

Do, T.V., Jorstad, I. (2005). A service-oriented architecture framework for mobile services. *Proceedings of Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/E-Learning on Telecommunications Workshop AICT/SAPIR/ ELETE*, pp. 65 – 70, ISBN 0-7695-2388-9, Lisbon, Portugal, July 17-20, 2005

FOAF, (2010). FOAF Vocabulary Specification. 26.09.2011, Available from http://xmlns.com/foaf/spec/

FOAF project, (2000). The Friend of a Friend (FOAF) project. 26.09.2011, Available from http://www.foaf-project.org/

Google, (2010). Google's Semantic Web Push: Rich Snippets Usage Growing. 26.09.2011, Available from http://www.readwriteweb.com/archives/google_semantic_web_push_rich_snippets_usage_grow.php

Gutheim, P. (2011). An ontology-based context inference service for mobile applications in next-generation networks. *IEEE Communications Magazine*, Vol. 49, No. 1, (Jan. 2011), pp. 60 – 66, ISSN 0163-6804

IST SPICE project, (2008). 26.09.2011, Available from http://www.ist-spice.org/

Khan, A., Asghar, S. and Fong, S. (2011). Framework of integrated Semantic Web Services and Ontology Development for Telecommunication Industry. *Journal of Emerging Technologies in Web Intelligence*, Vol. 3, No. 2, (May 2011), pp.110-119, ISSN 1798-0461

Kolberg, M., Merabti, M. & Moyer, S. (2010). Consumer communication applications drive integration and convergence. *IEEE Communications Magazine*, Vol.48, No. 12, (December 2010), pp. 24 - 24, ISSN 0163-6804

Niazi, R., Mahmoud Q.H. (2009). An Ontology-Based Framework for Discovering Mobile Services. *Proceedings of CNSR 2009 Seventh Annual Communication Networks and Services Research Conference*, pp.178-184, ISBN 978-0-7695-3649-1, Moncton, New Brunswick, Canada, May 11-13, 2009

Li, X.F., Peng, H., Li, Y. & Qiao, X.Q. (2010). Research on Constructing Telecom Services Domain Ontology. *Proceedings of 2010 International Conference on Computational Intelligence and Software Engineering (CiSE),* pp.1-4, ISBN 978-1-4244-5391-7, Wuhan, China, Dec. 10-12, 2010

Linked Data, (2007). Linked Data - Connect Distributed Data across the Web. 26.09.2011, Available from http://linkeddata.org/

McIlraith, S.A., Son, T.C. & Zeng H.L. (2001). Semantic Web services. *IEEE Intelligent Systems,* Vol. 16, No.2, (Mar.-Apr. 2001), pp. 46 – 53, ISSN 1541-1672

Miller, J. & Mukerji, J. (2003). MDA Guide V1.0.1. 26.09.2011, Available from http://www.omg.org/cgi-bin/doc?omg/03-06-01

Moerdijk, A.-J. & Klostermann, L. (2003). Opening the Networks with Parlay/OSA: Standards and Aspects Behind the APIs, *IEEE Network*, vol.17, no.3, (May-June 2003) pp: 58-64, ISSN 0890-8044

OMA, (2001). User Agent Profile. 26.09.2011, Available from http://www.openmobilealliance.org/tech/affiliates/wap/wap-248-uaprof-20011020-a.pdf

OMG. (2003). Model Driven Architecture. 26.09.2011, Available from http://www.omg.org/mda/

OMG. (2005). Unified Modeling Language. 26.09.2011, Available from http://www.omg.org/spec/UML/

OMG. (2005). XML Metadata Interchange. 26.09.2011, Available from http://www.omg.org/spec/XMI/

OMG. (2006). Meta Object Facility. 26.09.2011, Available from http://www.omg.org/mof/

OMG. (2008). Meta Object Facility (MOF) 2.0 Query/View/Transformation (QVT). 26.09.2011, Available from http://www.omg.org/spec/QVT/index.htm

OMG. (2009). Ontology Definition Metamodel (ODM). 26.09.2011, Available from http://www.omg.org/spec/ODM/1.0/

Park, K.L., Yoon, U.H. & Kim S.D. (2009). Personalized Service Discovery in Ubiquitous Computing Environments. *IEEE Pervasive Computing*, Vol.8, No.1, (Jan.-March 2009), pp.58 – 65, ISSN 1536-1268

Qiao, X.Q., Li X.F. and You, T. (2008). A Semantic Description Approach for Telecommunications Network Capability Services. *Proceedings of 11th Asia-Pacific Network Operations and Management Symposium,* pp.334-343, ISBN 978-3-540-88622-8, Beijing, China, October 22-24, 2008

Qiao, X.Q., Li, X.F., You, T., Sun, L.H. (2008). Semantic Telecommunications Network Capability Services. *Proceedings of the 3rd Asian Semantic Web Conference (ASWC 2008)*. pp. 508–523, ISBN 978-3-540-89703-3, Bangkok, Thailand, Dec. 8-11 2008

Rój, M. (2008). Recommendations for telecommunications services ontology, In: *IST SIMS project deliverables*, 26.09.2011, Available from http://www.ist-sims.org.

Selic, B. (2003). The pragmatics of model-driven development. *IEEE Software,* Vol. 20, No. 5, (Sept.-Oct. 2003), pp. 19 – 25, ISSN 0740-7459

Stanford Protégé team. (2004). Protégé. 26.09.2011, Available from http://protege.stanford.edu/

Su X.M., Alapnes, S. and Shiaa M.M. (2009). Mobile Ontology: Its Creation and Its Usage, In: *Constructing Ambient Intelligence*, H. Gerhauser, J. Hupp, C. Efstratiou and J. Heppner (Eds.), pp.75-79, Springer-Verlag Berlin Heidelberg, ISBN 3-642-10606-4, Germany

Twitter, (2010). Semantics, tagging and Twitter. 26.09.2011, Available from http://ml.sun.ac.za/2010/04/23/semantics-tagging-and-twitter/

Veijalainen, J. (2007) Developing Mobile Ontologies; Who, Why, Where, and How? *Proceedings of 2007 International Conference on Mobile Data Management*, pp. 398 – 401, ISBN 1-4244-1241-2, Mannheim, Germany, May 1-1, 2007

Veijalainen, J. (2008). Mobile Ontologies: Concept, Development, Usage, and Business Potential. *International Journal on Semantic Web and Information Systems (IJSWIS)*, Vol. 4, No. 1, (Sep. 2008), pp.20-34, ISSN: 1552-6283

Villalonga, C., Strohbach, M., Snoeck, N., Sutterer, M., Belaunde, M., Kovacs, E., Zhdanova, A.V., Goix, L.W., Droegehorn, O. (2009). Mobile Ontology: Towards a Standardized Semantic Model for the Mobile Domain. *Proceedings of the 1st International Workshop on Telecom Service Oriented Architectures (TSOA 2007) at the 5th International Conference on Service-Oriented Computing*, pp. 248-257, ISBN, 978-3-540-93850-7. Vienna, Austria, September 17, 2007

Vitvar, T., Viskova, J. (2005). Semantic-enabled Integration of Voice and Data Services: Telecommunication Use Case, *Proceedings of 2005 IEEE European Conference on Web Services (ECOWS2005)*, pp. 138-151, ISBN 0-7695-2484-2, Vaxjo, Sweden, November14-16, 2005

W3C. (2004). OWL Web Ontology Language Overview. 26.09.2011, Available from http://www.w3.org/TR/owl-features/

W3C. (2004). Resource Description Framework (RDF). 26.09.2011, Available from http://www.w3.org/RDF/

W3C, (2007). Composite Capabilities/Preference Profiles: Structure and Vocabularies 2.0. 26.09.2011, Available from http://www.w3.org/Mobile/CCPP/

Zander, S., Schandl, B. (2011). Semantic Web-enhanced Context-aware Computing in Mobile Systems: Principles and Application. In: *Mobile Computing Techniques in Emerging Markets: Systems, Applications and Services*, Kumar, A.V. Senthil (Eds.), Retrieved from http://eprints.cs.univie.ac.at/2870/

Zhu, J.W., Li, B., Wang, F., Wang, S.CH. (2010). An Overview of CRS4MO Project: Construction and Retrieval System for Mobile Ontology. *Journal of Computational Information Systems*, Vol.6, No.6, (June, 2010), pp. 2009-2015, ISSN 1553-9105

9

Quantum Secure Telecommunication Systems

Oleksandr Korchenko[1], Petro Vorobiyenko[2],
Maksym Lutskiy[1], Yevhen Vasiliu[2] and Sergiy Gnatyuk[1]
[1]*National Aviation University*
[2]*Odessa National Academy of Telecommunication named after O.S. Popov*
Ukraine

Our scientific field is still in its embryonic stage. It's great that we haven't been around for two thousands years. We are still at a stage where very, very important results occur in front of our eyes
Michael Rabin

1. Introduction

Today there is virtually no area where information technology (IT) is not used in some way. Computers support banking systems, control the work of nuclear power plants, and control aircraft, satellites and spacecraft. The high level of automation therefore depends on the security level of IT.

The main features of information security are confidentiality, integrity and availability. Only providing these all gives availability for development secure telecommunication systems. *Confidentiality* is the basic feature of information security, which ensures that information is accessible only to authorized users who have an access. *Integrity* is the basic feature of information security indicating its property to resist unauthorized modification. *Availability* is the basic feature of information security that indicates accessible and usable upon demand by an authorized entity.

One of the most effective ways to ensure confidentiality and data integrity during transmission is cryptographic systems. The purpose of such systems is to provide key distribution, authentication, legitimate users authorisation, and encryption. *Key distribution is one of the most important problems of cryptography*. This problem can be solved with the help of (SECOQC White Paper on Quantum Key Distribution and Cryptography, 2007; Korchenko et al., 2010a):

- *Classical information-theoretic schemes* (requires channel with noise; efficiency is very low, 1–5%).
- *Classical public-key cryptography schemes* (Diffie-Hellman scheme, digital envelope scheme; it has computational security).

- *Classical computationally secure symmetric-key cryptographic schemes* (requires a pre-installed key on both sides and can be used only as scheme for increase in key size but not as key distribution scheme).
- *Quantum key distribution* (provides information-theoretic security; it can also be used as a scheme for increase in key length).
- *Trusted Couriers Key Distribution* (it has a high price and is dependent on the human factor).

In recent years, quantum cryptography (QC) has attracted considerable interest. Quantum key distribution (QKD) (Bennett, 1992; Bennett et al., 1992; Bennett et al., 1995; Bennett & Brassard, 1984; Bouwmeester et al., 2000; Gisin et al., 2002; Lütkenhaus & Shields, 2009; Scarani et al., 2009; Vasiliu & Vorobiyenko 2006; Williams, 2011) plays a dominant role in QC. The overwhelming majority of theoretic and practical research projects in QC are related to the development of QKD protocols. The number of different quantum technologies is increasing, but there is no comprehensive information about classification of these technologies in scientific literature (there are only a few works concerning different classifications of QKD protocols, for example (Gisin et al., 2002; Scarani, et al., 2009)). This makes it difficult to estimate the level of the latest achievements and does not allow using quantum technologies with full efficiency. The main purpose of this chapter is the systematisation and classification of up-to-date effective quantum technologies of data (transmitted via telecommunication channels) security, analysis of their strengths and weaknesses, prospects and difficulties of implementation in telecommunication systems.

The first of all *quantum technologies of information security* consist of (Korchenko et al., 2010b):

- Quantum key distribution.
- Quantum secure direct communication.
- Quantum steganography.
- Quantum secret sharing.
- Quantum stream cipher.
- Quantum digital signature, etc.

The theoretical basis of quantum cryptography is stated in set of books and review papers (see e.g. Bouwmeester et al., 2000; Gisin et al., 2002; Hayashi, 2006; Imre & Balazs, 2005; Kollmitzer & Pivk, 2010; Lomonaco, 1998; Nielsen & Chuang, 2000; Schumacher & Westmoreland, 2010; Vedral, 2006; Williams, 2011).

2. Main approaches to quantum secure telecommunication systems construction

2.1 Quantum key distribution

QKD includes the following protocols: protocols using single (non-entangled) qubits (two-level quantum systems) and qudits (d-level quantum systems, d>2) (Bennett, 1992; Bennett et al., 1992; Bourennane et al., 2002; Bruss & Macchiavello, 2002; Cerf et al., 2002; Gnatyuk et al., 2009); protocols using phase coding (Bennett, 1992); protocols using entangled states (Ekert, 1991; Durt et al., 2004); decoy states protocols (Brassard et al., 2000; Liu et al., 2010; Peng et al., 2007; Yin et al., 2008; Zhao et al., 2006a, 2006b); and some

other protocols (Bradler, 2005; Lütkenhaus & Shields, 2009; Navascués & Acín, 2005; Pirandola et al., 2008).

The main task of QKD protocols is encryption key generation and distribution between two users connecting via quantum and classical channels (Gisin et al., 2002). In 1984 Ch. Bennett from IBM and G. Brassard from Montreal University introduced the first QKD protocol (Bennett & Brassard, 1984), which has become an alternative solution for the problem of key distribution. This protocol is called *BB84* (Bouwmeester et al., 2000) and it refers to QKD protocols using single qubits. The states of these qubits are the polarisation states of single photons. The BB84 protocol uses four polarisation states of photons (0°, 45°, 90°, 135°). These states refer to two mutually unbiased bases. Error searching and correcting is performed using classical public channel, which need not be confidential but only authenticated. For the detection of intruder actions in the BB84 protocol, an error control procedure is used, and for providing unconditionally security a privacy amplification procedure is used (Bennett et al., 1995). The efficiency of the BB84 protocol equals 50%. Efficiency means the ratio of the photons number which are used for key generation to the general number of transmitted photons.

Six-state protocol requires the usage of four states, which are the same as in the BB84 protocol, and two additional directions of polarization: right circular and left circular (Bruss, 1998). Such changes decrease the amount of information, which can be intercepted. But on the other hand, the efficiency of the protocol decreases to 33%.

Next, the *4+2 protocol* is intermediate between the BB84 and B92 protocol (Huttner et al., 1995). There are four different states used in this protocol for encryption: "0" and "1" in two bases. States in each base are selected non-orthogonal. Moreover, states in different bases must also be pairwise non-orthogonal. This protocol has a higher information security level than the BB84 protocol, when weak coherent pulses, but not a single photon source, are used by sender (Huttner et al., 1995). But the efficiency of the 4+2 protocol is lower than efficiency of BB84 protocol.

In the *Goldenberg-Vaidman protocol* (Goldenberg & Vaidman, 1995), encryption of "0" and "1" is performed using two orthogonal states. Each of these two states is the superposition of two localised normalised wave packets. For protection against intercept-resend attack, packets are sent at random times.

A modified type of Goldenberg-Vaidman protocol is called the *Koashi-Imoto protocol* (Koashi & Imoto, 1997). This protocol does not use a random time for sending packets, but it uses an interferometer's non-symmetrisation (the light is broken in equal proportions between both long and short interferometer arms).

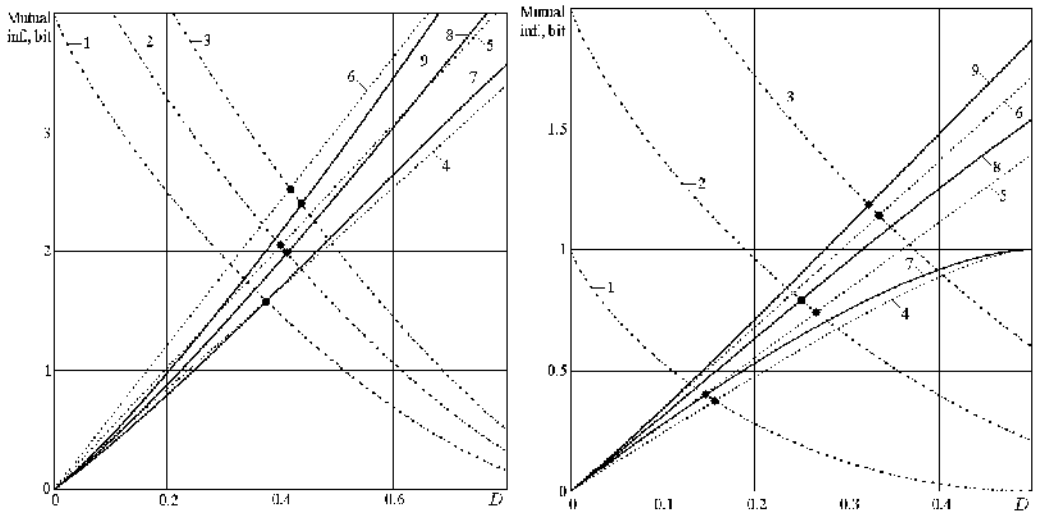
The measure of QKD protocol security is Shannon's mutual information between legitimate users (Alice and Bob) and an eavesdropper (Eve): $I_{AE}(D)$ and $I_{BE}(D)$, where D is error level which is created by eavesdropping. For most attacks on QKD protocols, $I_{AE}(D) = I_{BE}(D)$, we will therefore use $I_{AE}(D)$. The lower $I_{AE}(D)$ in the extended range of D is, the more secure the protocol is.

Six-state protocol and BB84 protocol were generalised in case of using d-level quantum systems — qudits instead qubits (Cerf et al., 2002). This allows increasing the information

capacity of protocols. We can transfer information using d-level quantum systems (which correspond to the usage of trits, quarts, etc.). It is important to notice that QKD protocols are intended for classical information (key) transfer via quantum channel.

The generalisation of BB84 protocol for qudits is called protocol using single qudits and two bases due to use of two mutually unbiased bases for the eavesdropping detection. Similarly, the generalisation of six-state protocol is called protocol using qudits and d+1 bases. These protocols' security against intercept-resend attack and non-coherent attack was investigated in a number of articles (see e.g. Cerf et al., 2002). Vasiliu & Mamedov have carried out a comparative analysis of the efficiency and security of different protocols using qudits on the basis of known formulas for mutual information (Vasiliu & Mamedov, 2008).

In fig. 1 dependences of $I_{AB}(D)$, $I_{AE}^{(d+1)}(D)$ and $I_{AE}^{(2)}(D)$ are presented, where $I_{AB}(D)$ is mutual information between Alice and Bob and $I_{AE}^{(d+1)}(D)$ and $I_{AE}^{(2)}(D)$ is mutual information between Alice and Eve for protocols using d+1 and two bases accordingly.



a) b)

Fig. 1. Mutual information for non-coherent attack. 1, 2, 3 — $I_{AB}(D)$ for d = 2, 4, 8 (a) and d = 16, 32, 64 (b); 4, 5, 6 — $I_{AE}^{(d+1)}(D)$ for d = 2, 4, 8 (a) and d = 16, 32, 64 (b); 7, 8, 9 — $I_{AE}^{(2)}(D)$ for d = 2, 4, 8 (a) and d = 16, 32, 64 (b).

In fig. 1 we can see that at low qudit dimension (up to d ~ 16) the protocol's security against non-coherent attack is higher when d+1 bases are used (when d = 2 it corresponds as noted above to greater security of six-state protocol than BB84 protocol). But the protocol's security is higher when two bases are used in the case of large d, while the difference in Eve's information (using d+1 or two bases) is not large in the work region of the protocol, i.e. in the region of Alice's and Bob's low error level. That's why that the number of bases used has little influence on the security of the protocol against non-coherent attack (at least for the qudit dimension up to d = 64). The crossing points of curves $I_{AB}(D)$ and $I_{AE}(D)$ correspond to boundary values D, up to which one's legitimate users can establish a secret

key by means of a privacy amplification procedure (even when eavesdropping occurs) (Bennett et al., 1995).

It is shown (Vasiliu & Mamedov, 2008) that the security of a protocol with qudits using two bases against intercept-resend attack is practically equal to the security of this protocol against non-coherent attack at any d. At the same time, the security of the protocol using d+1 bases against this attack is much higher. Intercept-resend attack is the weakest of all possible attacks on QKD protocols, but on the other hand, the efficiency of the protocol using d+1 bases rapidly decreases as d increases. A protocol with qudits using two bases therefore has higher security and efficiency than a protocol using d+1 bases.

Another type of QKD protocol is a *protocol using phase coding*: for example, the *B92 protocol* (Bennett, 1992) using strong reference pulses (Gisin et al., 2002). An eavesdropper can obtain more information about the encryption key in the B92 protocol than in the BB84 protocol for the given error level, however. Thus, the security of the B92 protocol is lower than the security of the BB84 protocol (Fuchs et al., 1997). The efficiency of the B92 protocol is 25%.

The *Ekert protocol (E91)* (Ekert, 1991) refers to QKD protocols using entangled states. Entangled pairs of qubits that are in a singlet state $\left|\psi^-\right\rangle = 1/\sqrt{2}\left(|0\rangle|1\rangle - |1\rangle|0\rangle\right)$ are used in this protocol. Qubit interception between Alice to Bob does not give Eve any information because no coded information is there. Information appears only after legitimate users make measurements and communicate via classical public authenticated channel (Ekert, 1991). But attacks with additional quantum systems (ancillas) are nevertheless possible on this protocol (Inamori et al., 2001).

Kaszlikowski et al. carried out the generalisation of the Ekert scheme for three-level quantum systems (Kaszlikowski et al., 2003) and Durt et al. carried out the generalisation of the Ekert scheme for d-level quantum systems (Durt et al., 2004): this increases the information capacity of the protocol a lot. Also the security of the protocol using entangled qudits is investigated (Durt et al., 2004). In the paper (Vasiliu & Mamedov, 2008), based on the results of (Durt et al., 2004), the security comparison of protocol using entangled qudits and protocols using single qudits (Cerf et al., 2002) against non-coherent attack is made. It was found that the security of these two kinds of protocols is almost identical. But the efficiency of the protocol using entangled qudits increases more slowly with the increasing dimension of qudits than the efficiency of the protocol using single qudits and two bases. Thus, from all contemporary QKD protocols using qudits, the most effective and secure against non-coherent attack is the protocol using single qudits and two bases (BB84 for qubits).

The aforementioned protocols with qubits are vulnerable to photon number splitting attack. This attack cannot be applied when the photon source emits exactly one photon. But there are still no such photon sources. Therefore, sources with Poisson distribution of photon number are used in practice. The part of pulses of this source has more than one photon. That is why Eve can intercept one photon from pulse (which contains two or more photons) and store it in quantum memory until Alice transfers Bob the sequence of bases used. Then Eve can measure stored states in correct basis and get the cryptographic key while

remaining invisible. It should be noted that there are more advanced strategies of photon number splitting attack which allow Bob to get the correct statistics of the photon number in pulses if Bob is controlling these statistics (Lutkenhaus & Jahma, 2002).

In practice for realisation of BB84 and six-state protocols weak coherent pulses with average photon number about 0,1 are used. This allows avoiding small probability of two- and multi-photon pulses, but this also considerably reduces the key rate.

The *SARG04 protocol* does not differ much from the original BB84 protocol (Branciard et al., 2005; Scarani et al., 2004; Scarani et al., 2009). The main difference does not refer to the "quantum" part of the protocol; it refers to the "classical" procedure of key sifting, which goes after quantum transfer. Such improvement allows increasing security against photon number splitting attack. The SARG04 protocol in practice has a higher key rate than the BB84 protocol (Branciard et al., 2005).

Another way of protecting against photon number splitting attack is the use of *decoy states QKD protocols* (Brassard et al., 2000; Peng et al., 2007; Rosenberg et al., 2007; Zhao et al., 2006), which are also advanced types of BB84 protocol. In such protocols, besides information signals Alice's source also emits additional pulses (decoys) in which the average photon number differs from the average photon number in the information signal. Eve's attack will modify the statistical characteristics of the decoy states and/or signal state and will be detected. As practical experiments have shown for these protocols (as for the SARG04 protocol), the key rate and practical length of the channel is bigger than for BB84 protocols (Peng et al., 2007; Rosenberg et al., 2007; Zhao et al., 2006). Nevertheless, it is necessary to notice that using these protocols, as well as the others considered above, it is also impossible without users pre-authentication to construct the complete high-grade solution of the problem of key distribution.

As a conclusion, after the analysis of the first and scale quantum method, we must sum up and highlight the following *advantages of QKD protocols:*

1. These protocols always allow eavesdropping to be detected because Eve's connection brings much more error level (compared with natural error level) to the quantum channel. The laws of quantum mechanics allow eavesdropping to be detected and the dependence between error level and intercepted information to be set. This allows applying privacy amplification procedure, which decreases the quantity of information about the key, which can be intercepted by Eve. Thus, QKD protocols have unconditional (information-theoretic) security.
2. The information-theoretic security of QKD allows using an absolutely secret key for further encryption using well-known classical symmetrical algorithms. Thus, the entire information security level increases. It is also possible to synthesize QKD protocols with Vernam cipher (one-time pad) which in complex with unconditionally secured authenticated schemes gives a totally secured system for transferring information.

The disadvantages of quantum key distribution protocols are:

1. A system based only on QKD protocols cannot serve as a complete solution for key distribution in open networks (additional tools for authentication are needed).

2. The limitation of quantum channel length which is caused by the fact that there is no possibility of amplification without quantum properties being lost. However, the technology of quantum repeaters could overcome this limitation in the near future (Sangouard et al., 2011).
3. Need for using weak coherent pulses instead of single photon pulses. This decreases the efficiency of protocol in practice. But this technology limitation might be defeated in the nearest future.
4. The data transfer rate decreases rapidly with the increase in the channel length.
5. Photon registration problem which leads to key rate decreasing in practice.
6. Photon depolarization in the quantum channel. This leads to errors during data transfer. Now the typical error level equals a few percent, which is much greater than the error level in classical telecommunication systems.
7. Difficulty of the practical realisation of QKD protocols for *d*-level quantum systems.
8. The high price of commercial QKD systems.

2.2 Quantum secure direct communication

The next method of information security based on quantum technologies is the usage of *quantum secure direct communication (QSDC) protocols* (Boström & Felbinger, 2002; Chuan et al., 2005; Cai, 2004; Cai & Li, 2004a; Cai & Li, 2004b; Deng et al., 2003; Vasiliu, 2011; Wang et al., 2005a, 2005b). The main feature of QSDC protocols is that there are no cryptographic transformations; thus, there is no key distribution problem in QSDC. In these protocols, a secret message is coded by qubits' (qudits') – quantum states, which are sent via quantum channel. QSDC protocols can be divided into several types:

- *Ping-pong protocol (and its enhanced variants)* (Boström & Felbinger, 2002; Cai & Li, 2004b; Chamoli & Bhandari, 2009; Gao et al., 2008; Ostermeyer & Walenta, 2008;Vasiliu & Nikolaenko, 2009; Vasiliu, 2011).
- *Protocols using block transfer of entangled qubits* (Deng et al., 2003; Chuan et al., 2005; Gao et al., 2005; Li et al., 2006; Lin et al., 2008; Xiu et al., 2009; Wang et al., 2005a, 2005b).
- *Protocols using single qubits* (Cai, 2004; Cai & Li, 2004a).
- *Protocols using entangled qudits* (Wang et al., 2005b; Vasiliu, 2011).

There are QSDC protocols for two parties and for multi-parties, e.g. broadcasting or when one user sends message to another under the control of a trusted third party.

Most contemporary protocols require a transfer of qubits by blocks (Chuan et al., 2005; Wang et al., 2005). This allows eavesdropping to be detected in the quantum channel before transfer of information. Thus, transfer will be terminated and Eve will not obtain any secret information. But for storing such blocks of qubits there is a need for a large amount of quantum memory. The technology of quantum memory is actively being developed, but it is still far from usage in common standard telecommunication equipment. So from the viewpoint of technical realisation, protocols using single qubits or their non-large groups (for one cycle of protocol) have an advantage. There are few such protocols and they have only asymptotic security, i.e. the attack will be detected with high probability, but Eve can obtain some part of information before detection. Thus, the problem of privacy amplification appears. In other words, new pre-processing methods of

transferring information are needed. Such methods should make intercepted information negligible.

One of the quantum secure direct communication protocols is the ping-pong protocol (Boström & Felbinger, 2002; Cai & Li, 2004b; Vasiliu, 2011), which does not require qubit transfer by blocks. In the first variant of this protocol, entangled pairs of qubits and two coding operations that allow the transmission of one bit of classical information for one cycle of the protocol are used (Boström & Felbinger, 2002). The usage of quantum superdense coding allows transmitting two bits for a cycle (Cai & Li, 2004b). The subsequent increase in the informational capacity of the protocol is possible by the usage instead of entangled pairs of qubits their triplets, quadruplets etc. in Greenberger-Horne-Zeilinger (GHZ) states (Vasiliu & Nikolaenko, 2009). The informational capacity of the ping-pong protocol with GHZ-states is equal to n bits on a cycle where n is the number of entangled qubits. Another way of increasing the informational capacity of ping-pong protocol is using entangled states of qudits. Thus, the corresponding protocol based on Bell's states of three-level quantum system (qutrit) pairs and superdense coding for qutrits is introduced (Wang et al., 2005; Vasiliu, 2011).

The advantages of QSDC protocols are a lack of secret key distribution, the possibility of data transfer between more than two parties, and the possibility of attack detection providing a high level of information security (up to information-theoretic security) for the protocols using block transfer. The main disadvantages are difficulty in practical realisation of protocols using entangled states (and especially protocols using entangled states for d-level quantum systems), slow transfer rate, the need for large capacity quantum memory for all parties (for protocols using block transfer of qubits), and the asymptotic security of the ping-pong protocol. Besides, QSDC protocols similarly to QKD protocols is vulnerable to man-in-the-middle attack, although such attack can be neutralized by using authentication of all messages, which are sent via the classical channel.

Asymptotic security of the ping-pong protocol (which is one of the simplest QSDC protocols from the technical viewpoint) can be amplified by using methods of classical cryptography. Security of several types of ping-pong protocols using qubits and qutrits against different attacks was investigated in series of papers (Boström & Felbinger, 2002; Cai, 2004; Vasiliu, 2011; Vasiliu & Nikolaenko, 2009; Zhang et al., 2005a).

The security of the ping-pong protocol using qubits against eavesdropping attack using ancilla states is investigated in (Boström & Felbinger, 2002; Chuan et al., 2005; Vasiliu & Nikolaenko, 2009).

Eve's information at attack with usage of auxiliary quantum systems (probes) on the ping-pong protocol with entangled n-qubit GHZ-states is defined by von Neumann entropy (Boström & Felbinger, 2002):

$$I_0 = S(\rho) \equiv -Tr\{\rho \log_2 \rho\} = -\sum_i \lambda_i \log_2 \lambda_i \tag{1}$$

where λ_i are the density matrix eigenvalues for the composite quantum system "transmitted qubits - Eve's probe".

For the protocol with Bell pairs and quantum superdence coding the density matrix ρ have size 4x4 and four nonzero eigenvalues:

$$\begin{aligned}\lambda_{1,2} &= \frac{1}{2}(p_1+p_2) \pm \frac{1}{2}\sqrt{(p_1+p_2)^2 - 16p_1p_2d(1-d)},\\ \lambda_{3,4} &= \frac{1}{2}(p_3+p_4) \pm \frac{1}{2}\sqrt{(p_3+p_4)^2 - 16p_3p_4d(1-d)}.\end{aligned} \tag{2}$$

For the protocol with GHZ-triplets a density matrix size is 16x16, and a number of nonzero eigenvalues is equal to eight. At symmetrical attack their kind is (Vasiliu & Nikolaenko, 2009):

$$\begin{aligned}\lambda_{1,2} &= \frac{1}{2}(p_1+p_2) \pm \frac{1}{2}\sqrt{(p_1+p_2)^2 - 16p_1p_2 \cdot \frac{2}{3}d\left(1-\frac{2}{3}d\right)},\\ \lambda_{7,8} &= \frac{1}{2}(p_7+p_8) \pm \frac{1}{2}\sqrt{(p_7+p_8)^2 - 16p_7p_8 \cdot \frac{2}{3}d\left(1-\frac{2}{3}d\right)}.\end{aligned} \tag{3}$$

For the protocol with *n*-qubit GHZ-states, the number of nonzero eigenvalues of density matrix is equal to 2^n, and their kind at symmetrical attack is (Vasiliu & Nikolaenko, 2009):

$$\begin{aligned}\lambda_{1,2} &= \frac{1}{2}(p_1+p_2) \pm \frac{1}{2}\sqrt{(p_1+p_2)^2 - 16p_1p_2 \cdot \frac{2^{n-2}}{2^{n-1}-1}d\left(1-\frac{2^{n-2}}{2^{n-1}-1}d\right)},\\ \lambda_{2^n-1,2^n} &= \frac{1}{2}\left(p_{2^n-1}+p_{2^n}\right) \pm \frac{1}{2}\sqrt{\left(p_{2^n-1}+p_{2^n}\right)^2 - 16p_{2^n-1}p_{2^n} \cdot \frac{2^{n-2}}{2^{n-1}-1}d\left(1-\frac{2^{n-2}}{2^{n-1}-1}d\right)},\end{aligned} \tag{4}$$

where *d* is probability of attack detection by legitimate users at one-time switching to control mode; p_i are frequencies of *n*-grams in the transmitted message.

The probability of that Eve will not be detected after *m* successful attacks and will gain information $I = mI_0$ is defined by the equation (Boström & Felbinger, 2002):

$$s(I,q,d) = \left(\frac{1-q}{1-q(1-d)}\right)^{I/I_0}, \tag{5}$$

where *q* is a probability of switching to control mode.

In fig. 2 dependences of $s(I,q,d)$ for several *n*, identical frequencies $p_i = 2^{-n}$, q = 0.5 and $d = d_{\max}$ are shown (Vasiliu & Nikolaenko, 2009). $d_{\max}$ is maximum probability of attack detection at one-time run of control mode, defined as

$$d_{\max} = 1 - \frac{1}{2^{n-1}}. \tag{6}$$

At $d = d_{\max}$ Eve gains the complete information about transmitted bits of the message. It is obvious from fig. 2 that the ping-pong protocol with many-qubit GHZ-states is asymptotically secure at any number n of qubits that are in entangled GHZ-states. A similar result for the ping-pong protocol using qutrit pairs is presented (Vasiliu, 2011).

A non-quantum method of security amplification for the ping-pong protocol is suggested in (Vasiliu & Nikolaenko, 2009; Korchenko et al., 2010c). Such method has been developed on the basis of a method of privacy amplification which is utilized in quantum key distribution protocols. In case of the ping-pong protocol this method can be some kind of analogy of the Hill cipher (Overbey et al., 2005).

Before the transmission Alice divides the binary message on l blocks of some fixed length r, we will designate these blocks as a_i (i=1,…l). Then Alice generates for each block separately random invertible binary matrix K_i of size $r \times r$ and multiplies these matrices by appropriate blocks of the message (multiplication is performed by modulo 2):

$$b_i = K_i a_i. \tag{7}$$

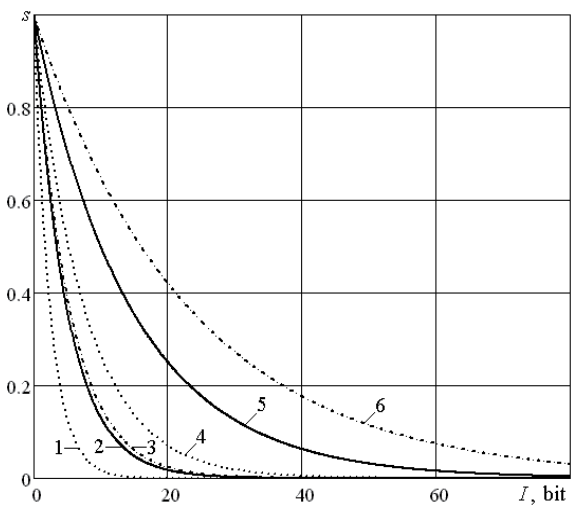


Fig. 2. Composite probability of attack non-detection s for the ping-pong protocol with many-qubit GHZ-states: n=2, original protocol (1); n=2, with superdense coding (2); n=3 (3); n=5 (4); n=10 (5); n=16 (6). I is Eve's information.

Blocks b_i are transmitted on the quantum channel with the use of the ping-pong protocol. Even if Eve, remained undetected, manages to intercept one (or more) from these blocks and without knowledge of used matrices K_i Eve won't be able to reconstruct source blocks a_i. To reach a sufficient security level the block length r and accordingly the size of matrices K_i should be selected so that Eve's undetection probability s after transmission of *one* block would be insignificant small. Matrices K_i are transmitted to Bob via usual (non-quantum) open authentic channel after the end of quantum transmission but only in the event when Alice and Bob were convinced lack of eavesdropping. Then Bob inverses the received matrices and having multiplied them on appropriate blocks b_i he gains an original message.

Let's mark that described procedure is not message enciphering, and can be named inverse hashing or hashing using two-way hash function, which role random invertible binary matrix acts.

It is necessary for each block to use individual matrix K_i which will allow to prevent cryptoanalytic attacks, similar to attacks to the Hill cipher, which are possible there at a multiple usage of one matrix for enciphering of several blocks (Eve could perform similar attack if she was able before a detection of her operations in the quantum channel to intercept several blocks, that are hashing with the same matrix). As matrices in this case are not a key and they can be transmitted on the open classical channel, the transmission of the necessary number of matrices is not a problem.

Necessary length r of blocks for hashing and accordingly necessary size $r \times r$ of hashing matrices should correspond to a requirement $r > I$, where I is the information which is gained by Eve. Thus, it is necessary for determination of r to calculate I at the given values of n, s, q and $d = d_{\max}$.

Let's accept $s(I, q, d) = 10^{-k}$, then:

$$I = \frac{-kI_0}{\lg\left(\frac{1-q}{1-q(1-d)}\right)}. \tag{8}$$

The calculated values of I are shown in tab. 1:

n	q = 0,5; $d = d_{\max}$	q = 0,5; $d = d_{\max}/2$	q = 0,25; $d = d_{\max}$	q = 0,25; $d = d_{\max}/2$
2	69	113	180	313
3	74	122	186	330
4	88	145	216	387
5	105	173	254	458
6	123	204	297	537
7	142	236	341	620
8	161	268	387	706
9	180	302	434	793
10	200	335	481	881
11	220	369	529	970
12	240	403	577	1059
13	260	437	625	1149
14	279	471	673	1238
15	299	505	721	1328
16	319	539	769	1417
17	339	573	817	1507
18	359	607	865	1597
19	379	641	913	1686
20	399	675	961	1776

Table 1. Eve's information I at attack on the ping - pong protocol with n-qubit GHZ-states at $s = 10^{-6}$ (bit).

Thus, after transfer of hashed block, the lengths of which are presented in tab. 1, the probability of attack non-detection will be equal to 10^{-6}; there is thus a very high probability that this attack will be detected. The main disadvantage of the ping-pong protocol, namely its asymptotic security against eavesdropping attack using ancilla states, is therefore removed.

There are some others attacks on the ping-pong protocol, e.g. attack which can be performed when the protocol is executed in quantum channel with noise (Zhang, 2005a) or Trojan horse attack (Gisin et al., 2002). But there are some counteraction methods to these attacks (Boström & Felbinger, 2008). Thus, we can say that the ping-pong protocol (the security of which is amplified using method described above) is the most prospective QSDC protocol from the viewpoint of the existing development level of the quantum technology of information processing.

2.3 Quantum steganography

Quantum steganography aims to hide the fact of information transferral similar to classical steganography. Most current models of quantum steganography systems use entangled states. For example, modified methods of entangled photon pair detection are used to hide the fact of information transfer in patent (Conti et al., 2004).

A simple quantum steganographic protocol (stegoprotocol) with using four qubit entangled Bell states:

$$\begin{aligned} \left|\phi^{+}\right\rangle &= \frac{1}{\sqrt{2}}\left(\left|0\right\rangle_1\left|0\right\rangle_2 + \left|1\right\rangle_1\left|1\right\rangle_2\right), \ \left|\phi^{-}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0\right\rangle_1\left|0\right\rangle_2 - \left|1\right\rangle_1\left|1\right\rangle_2\right), \\ \left|\psi^{+}\right\rangle &= \frac{1}{\sqrt{2}}\left(\left|0\right\rangle_1\left|1\right\rangle_2 + \left|1\right\rangle_1\left|0\right\rangle_2\right), \ \left|\psi^{-}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0\right\rangle_1\left|1\right\rangle_2 - \left|1\right\rangle_1\left|0\right\rangle_2\right), \end{aligned} \tag{9}$$

was proposed (Terhal et al., 2005). In this protocol n Bell states, including all four states (9) with equal probability is divided between two legitimate users (Alice and Bob) by third part (Trent). For all states the first qubit is sent to Alice and second to Bob. The secret bit is coded in the number of m singlet states $\left|\psi^{-}\right\rangle$ in the sequence of n states: even m represents "0" and odd represents "1". Alice and Bob perform local measurements each on own qubits and calculate the number of singlet states $\left|\psi^{-}\right\rangle$. That's why in this protocol Trent can secretly transmit information to Alice and Bob simultaneously.

Shaw & Brun proposed another one quantum stegoprotocol (Shaw & Brun, 2010). In this protocol the information qubit is hidden inside the error-correcting code. Thus, for intruder the qubits transmission via quantum channel looks like a normal quantum information transmission in the noise channel. For information qubit detection the receiver (Bob) must have a shared secret key with sender (Alice), which must be distributed before stegoprotocol starting. In the fig.3 the scheme of protocol proposed by Shaw & Brun is shown. Alice hides information qubit changing its places with qubit in her quantum codeword. She uses her secret key to determine which qubit in codeword must be replaced. Next, Alice uses key again to twirl (rotate) information qubit. This means that Alice uses one of the four single

qubit operators (Pauli operators) I, σ_x, σ_y or σ_z for this qubit by determining a concrete operation using two current key bits.

For the intruder who hasn't a key, this qubit likes qubit in maximal mixed state (the rotation can be interpreted as quantum Vernam cipher). In the next stage Alice uses random depolarization mistakes (using the same Pauli operators σ_x, σ_y or σ_z) to some part of others qubits of codeword for simulating some level of noise in quantum channel. Next, she sent a codeword to Bob. For correct untwirl operation Bob use the shared secret key and then he uses a key again to find information qubit.

The security of this protocol depends on the security of previous key distribution procedure. When key distribution has information-theoretic security, and using information qubit twirl (equivalent to quantum Vernam cipher) all scheme can have information-theoretic security. It is known the information-theoretic security is provided by QKD protocols. But if an intruder continuously monitors the channel for a long time and he has a precise channel characteristics, in the final he discovers that Alice transmits information to Bob on quantum stegoprotocol. In addition, using quantum measurements of transmitted qubit states, an intruder can cancel information transmitting (Denial of Service attack).

Thus, in the present three basis methods of quantum steganography are proposed:

1. Hiding in the quantum noise;
2. Hiding using quantum error-correcting codes;
3. Hiding in the data formats, protocols etc.

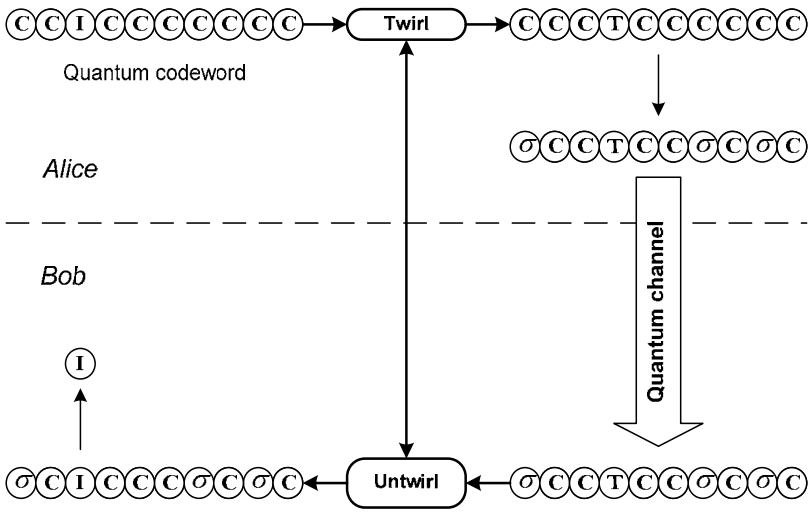


Fig. 3. The scheme of quantum stegoprotocol: C – qubit of codeword, I – information qubit, T – twirled information qubit, σ – qubit, to which Alice applies Pauli operator (qubit that simulate a noise).

The last method is the most promising direction of quantum steganography and also hiding using quantum error-correcting codes has some prospect in the future practice implementation.

It should be noted that theoretical research in quantum steganography has not reached the level of practical application yet, and it is very difficult to talk about the advantages and disadvantages of quantum steganography systems. Whether quantum steganography is superior to the classical one or not in practical use is still an open question (Imai & Hayashi, 2006).

2.4 Others technologies for quantum secure telecommunication systems construction

Quantum secret sharing (QSS). Most QSS protocols use properties of entangled states. The first QSS protocol was proposed by *Hillery, Buzek* and *Berthiaume* in 1998 (Hillery et al., 1998; Qin et al., 2007). This protocol uses GHZ-triplets (quadruplets) similar to some QSDC protocols. The sender shares his message between two (three) parties and only cooperation allows them to read this message. Semi-quantum secret sharing protocol using GHZ-triplets (quadruplets) was proposed by Li et al. (Li et al., 2009). In this protocol, users that receive a shared message have access to the quantum channel. But they are limited by some set of operation and are called "classical", meaning they are not able to prepare entangled states and perform any quantum operations or measurements. These users can measure qubits on a "classical" $\{|0\rangle,|1\rangle\}$ basis, reordering the qubits (via proper delay measurements), preparing (fresh) qubits in the classical basis, and sending or returning the qubits without disturbance. The sending party can perform any quantum operations. This protocol prevails over others QSS protocols in economic terms. Its equipment is cheaper because expensive devices for preparing and measuring (in GHZ-basis) many-qubit entangled states are not required. Semi-quantum secret sharing protocol exists in two variants: randomisation-based and measurement-resend protocols. Zhang et al. has been presented QSS using single qubits that are prepared in two mutually unbiased bases and transferred by blocks (Zhang et al., 2005b). Similar to the Hillery-Buzek-Berthiaume protocol, this allows sharing a message between two (or more) parties. The security improvement of this protocol against malicious acts of legitimate users is proposed (Deng et al., 2005). A similar protocol for multiparty secret sharing also is presented (Yan et al., 2008). QSS protocols are protected against external attackers and unfair actions of the protocol's parties. Both quantum and semi-quantum schemes allow detecting eavesdropping and do not require encryption unlike the classical secret-sharing schemes. The most significant imperfection of QSS protocols is the necessity for large quantum memory that is outside the capabilities of modern technologies today.

Quantum stream cipher (QSC) provides data encryption similar to classical stream cipher, but it uses quantum noise effect (Hirota et al., 2005) and can be used in optical telecommunication networks. QSC is based on the *Yuen-2000 protocol (Y-00, $\alpha\eta$ - scheme).* Information-theoretic security of the Y-00 protocol is ensured by randomisation (based on quantum noise) and additional computational schemes (Nair & Yuen, 2007; Yuen, 2001). In a number of papers (Corndorf et al., 2005; Hirota & Kurosawa, 2006; Nair & Yuen, 2007) the high encryption rate of the Y-00 protocol is demonstrated experimentally, and a security analysis on the Yuen-2000 protocol against the fast correlation attack, the typical attack on stream ciphers, is presented (Hirota & Kurosawa, 2006). The next advantage is better security compared with usual (classical) stream cipher. This is achieved by quantum noise

effect and by the impossibility of cloning quantum states (Wooters & Zurek, 1982). The complexity of practical implementation is the most important imperfection of QSC (Hirota & Kurosawa, 2006).

Quantum digital signature (QDS) can be implemented on the basis of protocols such as QDS protocols using single qubits (Wang et al., 2006) and QDS protocols using entangled states (authentic QDS based on quantum GHZ-correlations) (Wen & Liu, 2005). QDS is based on use of the quantum one-way function (Gottesman & Chuang, 2001). This function has better security than the classical one-way function, and it has information-theoretic security (its security does not depend on the power of the attacker's equipment). Quantum one-way function is defined by the following properties of quantum systems (Gottesman & Chuang, 2001):

1. Qubits can exist in superposition "0" and "1" unlike classical bits.
2. We can get only a limited quantity of classical information from quantum states according to the *Holevo theorem* (Holevo, 1977). Calculation and validation are not difficult but inverse calculation is impossible.

In the systems that use QDS, user identification and integrity of information is provided similar to classical digital signature (Gottesman & Chuang, 2001). The main advantages of QDS protocols are information-theoretic security and simplified key distribution system. The main disadvantage is the possibility to generate a limited number of public key copies and the leak of some quantities of information about incoming data of quantum one-way function (unlike the ideal classical one-way function) (Gottesman & Chuang, 2001).

Fig. 4 represents a general scheme of the methods of quantum secure telecommunication systems construction for their purposes and for using some quantum technologies.

2.5 Review of commercial quantum secure telecommunication systems

The world's first commercial quantum cryptography solution was *QPN Security Gateway (QPN-8505)* (QPN Security Gateway, 2011) proposed by *MagiQ Technologies (USA)*. This system (fig. 5 a) is a cost-effective information security solution for governmental and financial organisations. It proposes VPN protection using QKD (up to 100 256-bit keys per second, up to 140 km) and integrated encryption. The QPN-8505 system uses BB84, 3DES (NIST, 1999) and AES (NIST, 2001) protocols.

The Swiss company *Id Quantique* (Cerberis, 2011) offers a systems called *Clavis²* (fig. 5 b) and *Cerberis*. Clavis² uses a proprietary auto-compensating optical platform, which features outstanding stability and interference contrast, guaranteeing low quantum bit error rate. Secure key exchange becomes possible up to 100 km. This optical platform is well documented in scientific publications and has been extensively tested and characterized. Cerberis is a server with automatic creation and secret key exchange over a fibre channel (FC-1G, FC-2G and FC-4G). This system can transmit cryptographic keys up to 50 km and carries out 12 parallel cryptographic calculations. The latter substantially improves the system's performance. The Cerberis system uses AES (256-bits) for encryption and BB84 and SARG04 protocols for quantum key distribution. Main features:

- Future-proof security.

- Scalability: encryptors can be added when network grows.
- Versatility: encryptors for different protocols can be mixed.
- Cost-effectiveness: one quantum key server can distribute keys to several encryptors.

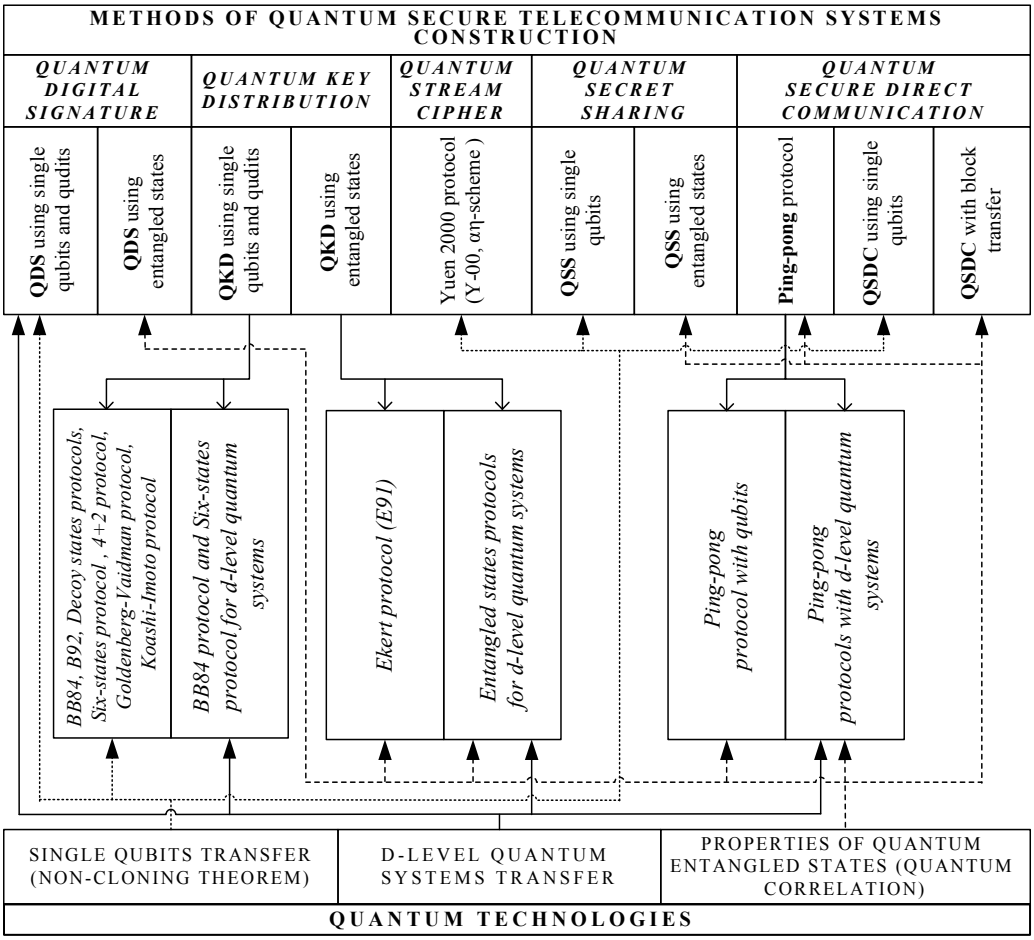


Fig. 4. Methods of quantum secure telecommunication systems construction.

Toshiba Research Europe Ltd (Great Britain) recently presented another QKD system named *Quantum Key Server* (QKS, 2011). This system (fig. 5 c) delivers digital keys for cryptographic applications on fibre optic based computer networks. Based on quantum cryptography it provides a failsafe method of distributing verifiably secret digital keys, with significant cost and key management advantages. The system provides world-leading performance. In particular, it allows key distribution over standard telecom fibre links exceeding 100 km in length and bit rates sufficient to generate 1 Megabit per second of key material over a distance of 50 km — sufficiently long for metropolitan coverage. Toshiba's system uses a

simple "one-way" architecture, in which the photons travel from sender to receiver. This design has been rigorously proven as secure from most types of eavesdropping attack. Toshiba has pioneered active stabilisation technology that allows the system to distribute key material continuously, even in the most challenging operating conditions, without any user intervention. This avoids the need for recalibration of the system due to temperature-induced changes in the fibre lengths. Initiation of the system is also managed automatically, allowing simple turn-key operation. It has been shown to work successfully in several network field trials. The system can be used for a wide range of cryptographic applications, e.g., encryption or authentication of sensitive documents, messages or transactions. A programming interface gives the user access to the key material.

a) b) c)

Fig. 5. Some commercial quantum secure telecommunication systems.

Another British company, *QinetiQ*, realised the world's first network using quantum cryptography—*Quantum Net (Qnet)* (Elliot et al., 2003; Hughes et al., 2002). The maximum length of telecommunication lines in this network is 120 km. Moreover, it is a very important fact that Qnet is the first QKD system using more than two servers. This system has six servers integrated to the Internet.

In addition the world's leading scientists are actively taking part in the implementation of projects such as *SECOQC (Secure Communication based on Quantum Cryptography)* (SECOQC White Paper on Quantum Key Distribution and Cryptography, 2007), *EQCSPOT (European Quantum Cryptography and Single Photon Technologies)* (Alekseev & Korneyko, 2007) and *SwissQuantum* (Swissquantum, 2011).

SECOQC is a project that aims to develop quantum cryptography network. The European Union decided in 2004 to invest € 11 million in the project as a way of circumventing espionage attempts by ECHELON (global intelligence gathering system, USA). This project combines people and organizations in Austria, Belgium, the United Kingdom, Canada, the Czech Republic, Denmark, France, Germany, Italy, Russia, Sweden and Switzerland. On October 8, 2008 SECOQC was launched in Vienna.

Following no-cloning theorem, QKD only can provide point-to-point (sometimes called "1:1") connection. So the number of links will increase $N(N-1)/2$ as N represents the number of nodes. If a node wants to participate into the QKD network, it will cause some issues like constructing quantum communication line. To overcome these issues, SECOQC was started. SECOQC network architecture (fig. 6) can by divided by two parts. Trusted private network and quantum network consisted with QBBs (Quantum Back Bone). Private network is conventional network with end-nodes and a QBB. QBB provides quantum

channel communication between QBBs. QBB is consisted with a number of QKD devices that are connected with other QKD devices in 1:1 connection. From this, SECOQC can provide easier registration of new end-node in QKD network, and quick recovery from threatening on quantum channel links.

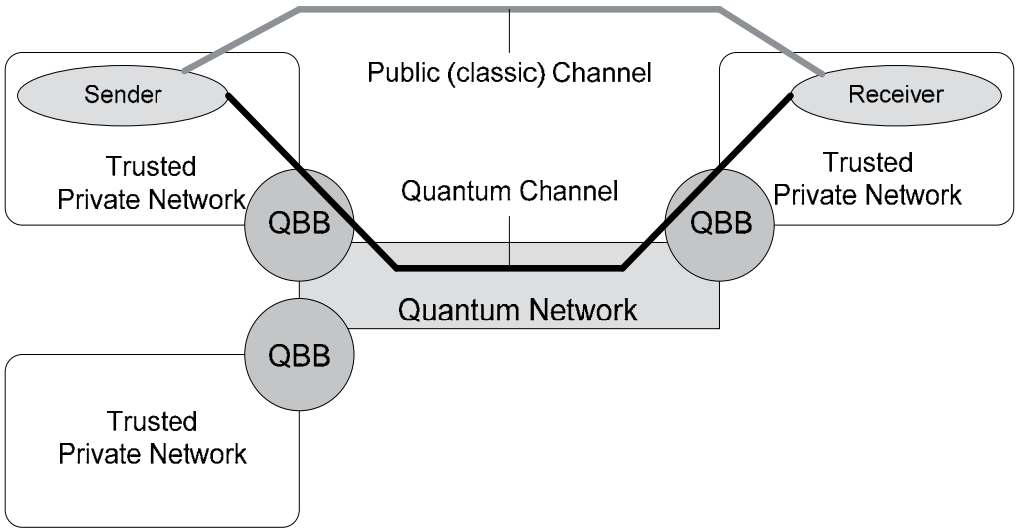


Fig. 6. Brief network architecture of SECOQC.

We also note that during the project SECOQC the seven most important QKD systems have been developed or refined (Kollmitzer & Pivk, 2010). Among these QKD systems are *Clavis*[2] and *Quantum Key Server* described above and also:

1. *The coherent one-way system (time-coding)* designed by GAP-Universite de Geneve and idQuantique realizes the novel distributed-phase-reference coherent one-way protocol.
2. *The entanglement-based QKD system* developed by an Austrian–Swedish consortium. The system uses the unique quantum mechanical property of entanglement for transferring the correlated measurements into a secret key.
3. *The free-space QKD system* developed by the group of H. Weinfurter from the University of Munich. It employs the BB84 protocol using polarization encoded attenuated laser pulses with photons of 850 nm wavelength. Decoy states are used to ensure key security even with faint pulses. The system is applicable to day and night operation using excessive filtering in order to suppress background light.
4. *The low-cost QKD system* was developed by John Rarity's team of the University of Bristol. The system can be applied for secure banking including consumer protection. The design philosophy is based on a future hand-held electronic credit card using free-space optics. A method is proposed to protect these transactions using the shared secret stored in a personal hand-held transmitter. Thereby Alice's module is integrated within a small device such as a mobile telephone, or personal digital

assistant, and Bob's module consists of a fixed device such as a bank asynchrone transfer mode.

The primary objective of EQCSPOT project is bringing quantum cryptography to the point of industrial application. Two secondary objectives exist to improve single photon technologies for wider applications in metrology, semiconductor characterisation, biosensing etc and to assess the practical use of future technologies for general quantum processors. The primary results will be in the tangible improvements in key distribution. The overall programme will be co-ordinated by British Defence Evaluation and Research Agency and the work will be divided into eight workparts with each workpart co-ordinated by one organisation. Three major workparts are dedicated to the development of the three main systems: NIR fibre, 1.3-1.55 μm fibre and free space key exchange. The other five are dedicated to networks, components and subsystems, software development, spin-off technologies and dissemination of results.

One of the key specificities of the SwissQuantum project is to aim at long-term demonstration of QKD and its applications. Although this is not the first quantum network to be deployed, it wills the first one to operate for months with real traffic. In this sense, the SwissQuantum network presents a major impetus for the QKD technology.

The SwissQuantum network consists of three layers:

- Quantum Layer. This layer performs Quantum Key Exchange.
- *Key Management Layer.* This layer manages the quantum keys in key servers and provides secure key storage, as well as advanced functions (key transfer and routing).
- *Application Layer.* In this layer, various cryptographic services use the keys distributed to provide secure communications.

There are many practical and theoretical research projects concerning the development of quantum technology in research institutes, laboratories and centres such as Institute for Quantum Optics and Quantum Information, Northwestern University, SmartQuantum, BBN Technologies of Cambridge, TREL, NEC, Mitsubishi Electric, ARS Seibersdorf Research and Los Alamos National Laboratory.

3. Conclusion

This chapter presents a classification and systematisation of modern quantum technology of information security. The characteristic of the basic directions of quantum cryptography from the point of view of the quantum technologies used is given. A qualitative analysis of the advantages and imperfections of concrete quantum protocols is made. Today the most developed direction of quantum secure telecommunication systems is QKD protocols. In research institutes, laboratories and centres, quantum cryptographic systems for secret key distribution for distant legitimate users are being developed. Most of the technologies used in these systems are patented in different countries (mainly in the U.S.A.). Such QKD systems can be combined with any classical cryptographic scheme, which provides information-theoretic security, and the entire cryptographic scheme will have information-theoretic security also. QKD protocols can generally provide higher information security level than appropriate classical schemes.

Other secure quantum technologies in practice have not been extended beyond laboratory experiments yet. But there are many theoretical cryptographic schemes that provide high information security level up to the information-theoretic security. QSDC protocols remove the secret key distribution problem because they do not use encryption. One of these is the ping-pong protocol and its improved versions. These protocols can provide high information security level of confidential data transmission using the existing level of technology with security amplification methods. Another category of QSDC is protocols with transfer qubits by blocks that have unconditional security, but these need a large quantum memory which is out of the capabilities of modern technologies today. It must be noticed that QSDC protocols are not suitable for the transfer of a high-speed flow of confidential data because there is low data transfer rate in the quantum channel. But when a high information security level is more important than transfer rate, QSDC protocols should find its application.

Quantum secret sharing protocols allow detecting eavesdropping and do not require data encryption. This is their main advantage over classical secret sharing schemes. Similarly, quantum stream cipher and quantum digital signature provide higher security level than classical schemes. Quantum digital signature has information-theoretic security because it uses quantum one-way function. However, practical implementation of these quantum technologies is also faced to some technological difficulties.

Thus, in recent years quantum technologies are rapidly developing and gradually taking their place among other means of information security. Their advantage is a high level of security and some properties, which classical means of information security do not have. One of these properties is the ability always to detect eavesdropping. Quantum technologies therefore represent an important step towards improving the security of telecommunication systems against cyber-terrorist attacks. But many theoretical and practical problems must be solved for wide practical use of quantum secure telecommunication systems.

4. Acknowledgment

Special thanks should be given to **Rector of National Aviation University (Kyiv, Ukraine) – Mykola Kulyk.** We would not have finished this chapter without his support.

5. References

Alekseev, D.A. & Korneyko, A.V. (2007). Practice reality of quantum cryptography key distribution systems, *Information Security*, No. 1, pp. 72–76.

Bennett, C. & Brassard, G. (1984). Quantum cryptography: public key distribution and coin tossing, *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing*. Bangalore, India, pp. 175–179.

Bennett, C. (1992). Quantum cryptography using any two non-orthogonal states, *Physical Review Letters*, Vol.68, No.21, pp. 3121–3124.

Bennett, C.; Bessette, F. & Brassard, G. (1992). Experimental Quantum Cryptography, *Journal of Cryptography*, Vol.5, No.1, pp. 3–28.

Bennett, C.; Brassard, G.; Crépeau, C. & Maurer, U. (1995). Generalized privacy amplification, *IEEE Transactions on Information Theory*, Vol.41, No.6, pp. 1915–1923.

Boström, K. & Felbinger, T. (2002). Deterministic secure direct communication using entanglement, *Physical Review Letters*, Vol.89, No.18, 187902.

Boström, K. & Felbinger, T. (2008). On the security of the ping-pong protocol, *Physics Letters A*, Vol.372, No.22, pp. 3953–3956.

Bourennane, M.; Karlsson, A. & Bjork, G. (2002). Quantum key distribution using multilevel encoding, *Quantum Communication, Computing, and Measurement 3*. N.Y.: Springer US, pp. 295–298.

Bouwmeester, D.; Ekert, A. & Zeilinger, A. (2000). *The Physics of Quantum Information. Quantum Cryptography, Quantum Teleportation, Quantum Computation*. Berlin: Springer-Verlag, 314 p.

Bradler K. (2005). Continuous variable private quantum channel, *Physical Review A*, Vol.72, No.4, 042313.

Branciard, C.; Gisin, N.; Kraus, B. & Scarani, V. (2005). Security of two quantum cryptography protocols using the same four qubit states, *Physical Review A*, Vol.72, No.3, 032301.

Brassard, G.; Lutkenhaus, N.; Mor, T. & Sanders, B. (2000). Limitations on practical quantum cryptography, *Physical Review Letters*, Vol.85, No.6, pp. 1330–1333.

Bruss, D. (1998). Optimal Eavesdropping in Quantum Cryptography with Six States, *Physical Review Letters*, Vol.81, No.14, pp. 3018–3021.

Bruss, D. & Macchiavello C. (2002). Optimal eavesdropping in cryptography with three-dimensional quantum states, *Physical Review Letters*, Vol.88, No.12, 127901.

Cai, Q.-Y. & Li, B.-W. (2004a). Deterministic Secure Communication Without Using Entanglement, *Chinese Physics Letters*, Vol.21 (4), pp. 601–603.

Cai, Q.-Y. & Li B.-W. (2004b). Improving the capacity of the Bostrom–Felbinger protocol, *Physical Review A*, Vol.69, No.5, 054301.

Cerberis. 01.10.2011, Available from: http://idquantique.com/products/cerberis.htm.

Cerf, N.J.; Bourennane, M.; Karlsson, A. & Gisin, N. (2002). Security of quantum key distribution using d-level systems, *Physical Review Letters*, Vol.88, No.12, 127902.

Chamoli, A. & Bhandari, C.M. (2009). Secure direct communication based on ping-pong protocol, *Quantum Information Processing*, Vol.8, No.4, pp. 347–356.

Chuan, W.; Fu Guo, D. & Gui Lu, L. (2005). Multi-step quantum secure direct communication using multi-particle Greenberg-Horne-Zeilinger state, *Optics Communications*, Vol.253, pp. 15–19.

Conti A.; Ralph, S.; Kenneth A. et al. *Patent* No 7539308 USA, H04K 1/00 (20060101). Quantum steganography, publ. 21.05.2004.

Corndorf, E., Liang, C. & Kanter, G.S. (2005). Quantum-noise randomized data encryption for wavelength-division-multiplexed fiber-optic networks, *Physical Review A*, Vol.71, No.6, 062326.

Deng, F.G.; Long, G.L. & Liu, X.S. (2003). Two-step quantum direct communication protocol using the Einstein–Podolsky–Rosen pair block. *Physical Review A*, 2003. Vol.68, No.4, 042317.

Deng, F. G.; Li, X. H.; Zhou, H. Y. & Zhang, Z. J. (2005). Improving the security of multiparty quantum secret sharing against Trojan horse attack, *Physical Review A*, Vol.72, No.4, 044302.

Desurvire, E. (2009). *Classical and Quantum Information Theory*. Cambridge: Cambridge University Press, 691 p.

Durt, T.; Kaszlikowski, D.; Chen, J.-L. & Kwek, L.C. (2004). Security of quantum key distributions with entangled qudits, *Physical Review A*, Vol.69, No.3, 032313.

Ekert, A. (1991). Quantum cryptography based on Bell's theorem, *Physical Review Letters*, Vol.67, No.6, pp. 661–663.

Elliot, C.; Pearson, D. & Troxel, G. (2003). Quantum Cryptography in Practice, *arXiv:quant-ph/0307049.*

Fuchs, C.; Gisin, N.; Griffits, R. et al. (1997). Optimal Eavesdropping in Quantum Cryptography. Information Bound and Optimal Strategy, *Physical Review A*, Vol.56, No.2, pp. 1163–1172.

Gao, T.; Yan, F.L. & Wang, Z.X. (2005). Deterministic secure direct communication using GHZ-states and swapping quantum entanglement. *Journal of Physics A: Mathematical and Theoretical*, Vol. 38, No.25, pp. 5761–5770.

Gao, F.; Guo, F.Zh.; Wen, Q.Y. & Zhu, F.Ch. (2008). Comparing the efficiencies of different detect strategies in the ping-pong protocol, *Science in China, Series G: Physics, Mechanics & Astronomy*, Vol.51, No.12. pp. 1853–1860.

Gisin, N.; Ribordy, G.; Tittel, W. & Zbinden, H. (2002). Quantum cryptography, *Review of Modern Physics*, Vol.74, pp. 145–195.

Gnatyuk, S.O.; Kinzeryavyy, V.M.; Korchenko, O.G. & Patsira, Ye.V. (2009). *Patent* No 43779 UA, MPK H04L 9/08. System for cryptographic key transfer, 25.08.2009.

Goldenberg, L. & Vaidman, L. (1995). Quantum Cryptography Based On Orthogonal States, *Physical Review Letters*, Vol.75, No.7, pp. 1239–1243.

Gottesman, D. & Chuang, I. (2001). Quantum digital signatures, *arXiv:quant-ph/0105032v2*.

Hayashi, M. (2006). *Quantum information. An introduction.* Berlin, Heidelberg, New York: Springer, 430 p.

Hillery, M.; Buzek, V. & Berthiaume, A. (1999). Quantum secret sharing, *Physical Review A*, Vol.59, No.3, pp. 1829–1834.

Hirota, O. & Kurosawa, K. (2006). An immunity against correlation attack on quantum stream cipher by Yuen 2000 protocol, *arXiv:quant-ph/0604036v1.*

Hirota, O.; Sohma, M.; Fuse, M. & Kato, K. (2005). Quantum stream cipher by the Yuen 2000 protocol: Design and experiment by an intensity-modulation scheme, *Physical Review A*, Vol.72, No.2, 022335.

Holevo, A.S. (1977). Problems in the mathematical theory of quantum communication channels, *Report of Mathematical Physics*, Vol.12, No.2, pp. 273–278.

Hughes, R.; Nordholt, J.; Derkacs, D. & Peterson, C. (2002). Practical free-space quantum key distribution over 10 km in daylight and at night, *New Journal of Physics*, Vol.4, 43 p.

Huttner, B.; Imoto, N.; Gisin, N. & Mor, T. (1995). Quantum Cryptography with Coherent States, *Physical Review A*, Vol.51, No.3, pp. 1863–1869.

Imai, H. & Hayashi, M. (2006). *Quantum Computation and Information. From Theory to Experiment*. Berlin: Springer-Verlag, Heidelberg, 235 p.

Imre, S. & Balazs, F. (2005). *Quantum Computing and Communications: An Engineering Approach,* John Wiley & Sons Ltd, 304 p.

Inamori, H.; Rallan, L. & Vedral, V. (2001). Security of EPR-based quantum cryptography against incoherent symmetric attacks, *Journal of Physics A*, Vol.34, No.35, pp. 6913–6918.

Kaszlikowski, D.; Christandl, M. et al. (2003). Quantum cryptography based on qutrit Bell inequalities, *Physical Review A*, Vol.67, No.1, 012310.

Koashi, M. & Imoto, N. (1997). Quantum Cryptography Based on Split Transmission of One-Bit Information in Two Steps, *Physical Review Letters*, Vol.79, No.12, pp. 2383–2386.

Kollmitzer, C. & Pivk, M. (2010). Applied Quantum Cryptography, *Lecture Notes in Physics 797*. Berlin, Heidelberg: Springer, 214 p.

Korchenko, O.G.; Vasiliu, Ye.V. & Gnatyuk, S.O. (2010a). Modern quantum technologies of information security against cyber-terrorist attacks, *Aviation.* Vilnius: Technika, Vol.14, No.2, pp. 58–69.

Korchenko, O.G.; Vasiliu, Ye.V. & Gnatyuk, S.O. (2010b). Modern directions of quantum cryptography, "*AVIATION IN THE XXI-st CENTURY*" – "*Safety in Aviation and Space Technologies*": *IV World Congress: Proceedings* (September 21–23, 2010), Kyiv, NAU, pp. 17.1–17.4.

Korchenko, O.G.; Vasiliu, Ye.V.; Nikolaenko, S.V. & Gnatyuk, S.O. (2010c). Security amplification of the ping-pong protocol with many-qubit Greenberger-Horne-Zeilinger states, *XIII International Conference on Quantum Optics and Quantum Information (ICQOQI'2010):* Book of abstracts (May 28 – June 1, 2010), pp. 58–59.

Li, Q.; Chan, W. H. & Long, D-Y. (2009). Semi-quantum secret sharing using entangled states, *arXiv:quant-ph/0906.1866v3*.

Li, X.H.; Deng, F.G. & Zhou, H.Y. (2006). Improving the security of secure direct communication based on the secret transmitting order of particles. *Physical Review A*, Vol.74, No.5, 054302.

Lin, S.; Wen, Q.Y.; Gao, F. & Zhu F.C. (2008). Quantum secure direct communication with chi-type entangled states, *Physical Review A*, Vol.78, No.6, 064304.

Liu, Y.; Chen, T.-Y.; Wang, J. et al. (2010). Decoy-state quantum key distribution with polarized photons over 200 km, *Optics Express*, Vol. 18, Issue 8, pp. 8587-8594.

Lomonaco, S.J. (1998). A Quick Glance at Quantum Cryptography, *arXiv:quant-ph/*9811056.

Lütkenhaus, N. & Jahma, M. (2002). Quantum key distribution with realistic states: photon-number statistics in the photon-number splitting attack, *New Journal of Physics*, Vol.4, pp. 44.1–44.9.

Lütkenhaus, N. & Shields, A. (2009). Focus on Quantum Cryptography: Theory and Practice, *New Journal of Physics*, Vol.11, No.4, 045005.

Nair, R. & Yuen, H. (2007). On the Security of the Y-00 (AlphaEta) Direct Encryption Protocol, *arXiv:quant-ph/0702093v2.*

Navascués, M. & Acín, A. (2005). Security Bounds for Continuous Variables Quantum Key Distribution, *Physical Review Letters*, Vol.94, No.2, 020505.

Nielsen, M.A. & Chuang, I.L. (2000). *Quantum Computation and Quantum Information*. Cambridge: Cambridge University Press, 676 p.

NIST. "FIPS-197: Advanced Encryption Standard." (2001). 01.10.2011, Available from: <http://csrc.nist.gov/publications/fips>.

NIST. "FIPS-46-3: Data Encryption Standard." (1999). 01.10.2011, Available from: <http://csrc.nist.gov/publications/fips>.

Ostermeyer, M. & Walenta N. (2008). On the implementation of a deterministic secure coding protocol using polarization entangled photons, *Optics Communications*, Vol. 281, No.17, pp. 4540–4544.

Overbey, J; Traves, W. & Wojdylo J. (2005). On the keyspace of the Hill cipher, *Cryptologia*, Vol.29, No.1, pp. 59–72.

Peng, C.-Z.; Zhang, J.; Yang, D. et al. (2007). Experimental long-distance decoy-state quantum key distribution based on polarization encoding, *Physical Review Letters*, Vol.98, No.1, 010505.

Pirandola, S.; Mancini, S.; Lloyd, S. & Braunstein S. (2008). Continuous-variable quantum cryptography using two-way quantum communication, *Nature Physics*, Vol.4, No.9, pp. 726–730.

Qin, S.-J.; Gao, F. & Zhu, F.-Ch. (2007). Cryptanalysis of the Hillery-Buzek-Berthiaume quantum secret-sharing protocol, *Physical Review A*, Vol.76, No.6, 062324.

QKS. Toshiba Research Europe Ltd. 01.10.2011, Available from: <http://www.toshiba-europe.com/research/crl/QIG/quantumkeyserver.html>.

QPN Security Gateway (QPN–8505). 01.10.2011, Available from: <http://www.magiqtech.com/MagiQ/Products.html>.

Rosenberg, D. et al. (2007). Long-distance decoy-state quantum key distribution in optical fiber, *Physical. Review Letters*, Vol.98, No.1, 010503.

Sangouard, N.; Simon, C.; de Riedmatten, H. & Gisin, N. (2011). Quantum repeaters based on atomic ensembles and linear optics, *Review of Modern Physics,* Vol.83, pp. 33–34.

Scarani, V.; Acin, A.; Ribordy, G. & Gisin, N. (2004). Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations, *Physical Review Letters*, Vol.92, No.5, 057901.

Scarani, V.; Bechmann-Pasquinucci, H.; Nicolas J. Cerf et al. (2009). The security of practical quantum key distribution, *Review of Modern Physics,* Vol.81, pp. 1301–1350.

SECOQC White Paper on Quantum Key Distribution and Cryptography. (2007). *arXiv:quant-ph/0701168v1*.

Shaw, B. & Brun, T. (2010). Quantum steganography, *arXiv:quant-ph/1006.1934v1.*

Schumacher, B. & Westmoreland, M. (2010). *Quantum Processes, Systems, and Information*. Cambridge: Cambridge University Press, 469 p.

Terhal, B.M.; DiVincenzo, D.P. & Leung, D.W. (2001). Hiding bits in Bell states, *Physical review letters,* Vol.86, issue 25, pp. 5807-5810.

Vasiliu, E.V. (2011). Non-coherent attack on the ping-pong protocol with completely entangled pairs of qutrits, *Quantum Information Processing,* Vol.10, No.2, pp. 189–202.

Vasiliu, E.V. & Nikolaenko, S.V. (2009). Synthesis if the secure system of direct message transfer based on the ping–pong protocol of quantum communication, *Scientific works of the Odessa national academy of telecommunications named after O.S. Popov*, No.1, pp. 83–91.

Vasiliu, E.V. & Mamedov, R.S. (2008). Comparative analysis of efficiency and resistance against not coherent attacks of quantum key distribution protocols with transfer of multidimensional quantum systems, *Scientific works of the Odessa national academy of telecommunications named after O.S. Popov*, No.2, pp. 20–27.

Vasiliu, E.V. & Vorobiyenko, P.P. (2006). The development problems and using prospects of quantum cryptographic systems, *Scientific works of the Odessa national academy of telecommunications named after O.S. Popov*, No.1, pp. 3–17.

Vedral, V. (2006). *Introduction to Quantum Information Science*. Oxford University Press Inc., New York, 183 p.

Wang, Ch.; Deng, F.G. & Long G.L. (2005a). Multi – step quantum secure direct communication using multi – particle Greenberger – Horne – Zeilinger state, *Optics Communications*, Vol. 253, No.1, pp. 15–20.

Wang, Ch. et al. (2005b). Quantum secure direct communication with high dimension quantum superdense coding, *Physical Review A*, Vol.71, No.4, 044305.

Wang, J.; Zhang, Q. & Tang, C. (2006). Quantum signature scheme with single photons, *Optoelectronics Letters*, Vol.2, No.3, pp. 209–212.

Wen, X.-J. & Liu, Y. (2005). Quantum Signature Protocol without the Trusted Third Party, *arXiv:quant-ph/0509129v2*.

Williams, C.P. (2011). *Explorations in quantum computing, 2nd edition*. Springer-Verlag London Limited, 717 p.

Wooters, W.K. & Zurek, W.H. (1982). A single quantum cannot be cloned, *Nature*, Vol. 299, p. 802.

Xiu, X.-M.; Dong, L.; Gao, Y.-J. & Chi F. (2009). Quantum Secure Direct Communication with Four-Particle Genuine Entangled State and Dense Coding, *Communication in Theoretical Physics*, Vol.52, No.1, pp. 60–62.

Yan, F.-L.; Gao, T. & Li, Yu.-Ch. (2008). Quantum secret sharing protocol between multiparty and multiparty with single photons and unitary transformations, *Chinese Physics Letters*, Vol.25, No.4, pp. 1187–1190.

Yin, Z.-Q.; Zhao, Y.-B.; Zhou Z.-W. et al. (2008). Decoy states for quantum key distribution based on decoherence-free subspaces, *Physical Review A*, Vol.77, No.6, 062326.

Yuen, H.P. (2001). *In Proceedings of QCMC'00*, Capri, edited by P. Tombesi and O. Hirota New York: Plenum Press, p. 163.

Zhang, Zh.-J.; Li, Y. & Man, Zh.-X. (2005a). Improved Wojcik's eavesdropping attack on ping-pong protocol without eavesdropping-induced channel loss, *Physics Letters A*, Vol.341, No.5–6, pp. 385–389.

Zhang, Zh.-J.; Li, Y. & Man, Zh.-X. (2005b). Multiparty quantum secret sharing, *Physical Review A*, Vol.71, No.4, 044301.

Zhao, Y.; Qi, B.; Ma, X.; Lo, H.-K. & Qian, L. (2006a). Simulation and implementation of decoy state quantum key distribution over 60 km telecom fiber, *Proceedings of IEEE International Symposium on Information Theory*, pp. 2094–2098.

Zhao, Y.; Qi, B.; Ma, X.; Lo, H.-K. & Qian, L. (2006b). Experimental Quantum Key Distribution with Decoy States, *Physical Review Letters*, Vol.96, No.7, 070502.

10

Web-Based Laboratory Using Multitier Architecture

C. Guerra Torres and J. de León Morales
Facultad de Ingenieria Mecánica y Eléctrica
Universidad Autónoma de Nuevo León
México

1. Introduction

Actuality, Internet provides a convenient way to develop a new communication technology for several applications, for example remote laboratories. The remote access to complex and expensive laboratories offers a cost-effective and flexible means for distance learning, research and remote experimentation. In the literature, some works propose platforms based on the Internet in order to access experimental laboratories; nevertheless it is necessary that the platform provides a good architecture, clear methodology of operation, and it must facilitate the integration between hardware (HW) and software (SW) elements. In this work, we present a platform based on "multitier programming architecture" which allows the easy integration of HW and SW elements and offers several schemes of tele-presence: teleoperation, telecontrol and teleprogramming.

The remote access to complex and expensive laboratory equipment represents an appealing issue and great interest for research, learning education and industrial applications. The range potentially involved is very large, including among others, applications in all fields of engineering (Restivo et al., 2009; Wu et al., 2008).

It is well known that several experimental platforms are distributed in different laboratories in the world, and all of them are on-line accessible through the Internet. Since those laboratories require specific resources to enable a remote access, several solutions for harmonizing the necessary software and hardware have been proposed and described. Furthermore, due to their versatility, these platforms provide user services which allow the transmission of information in a simply way, besides being available to many people, having many multimedia resources.

The potentiality of remote laboratories (Gomez & Garcia, 2007) and the use of the Internet, as a channel of communication to reach the students at their homes, were soon recognized (Basigalup et al., 2006; Davoli et al., 2006; Callangan et al., 2005; Imbre & Spong, 2006; Rapuano & Soino, 2005).

Several works based on remote experimentation, which are used as excellent alternatives to access remote equipment, have been published (Costas et al., 2008).

Then, to solve the problem of testing engineering algorithms in real-time, we apply the advantages of the computer Network, computer communication and teleoperation. Furthermore, developing these new tools give the possibility to use these equipments for remote education.

In remote experimentation there exists several schemes based on the communication channel called **telepresence schemes**, some of them are: i) **teleoperation**, ii) **teleprogramming** and iii) **telecontrol**. In (Wang & James, 2005) some concepts are related with teleoperation. In other works, (Huijun et al., 2008) analyze the time-delay in the telecontrol systems, and (Cloosterman et al., 2009) studies the stability of the feedback systems with With Uncertain Time-Varying Delays. Others authors propose platforms only to move remote equipment, for example robots, (Wang & James, 2005). Finally, few works talking about the remote programming are published; see for instance (Costas et al., 2008).

However, for a remote laboratory to be functional, it must be capable of offering different schemes of telepresence. This can be easily understood from figure 1 which is an extension of the figure given in (Baccigalup et al., 2006). A comparison between different teaching methods, taking into account the teaching effectiveness, time and cost per students, is schematized in figure 1.

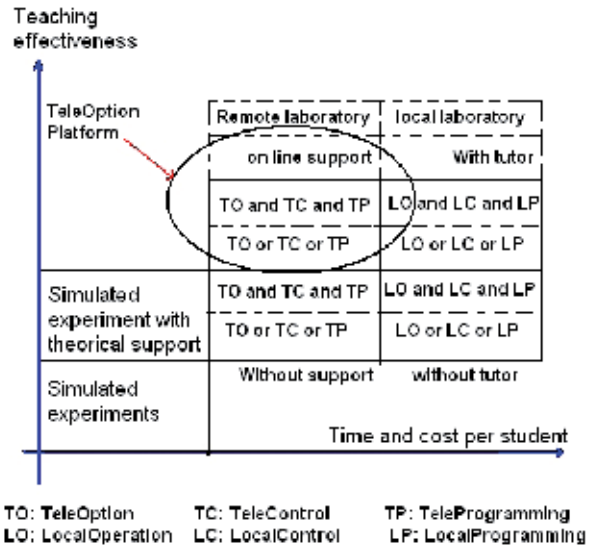


Fig. 1. Comparison between local and remote laboratories.

Contribution

Considering figure 1, the goal of this work is to introduce a platform called *Teleoptions*, which offers an alternative for remote laboratories, using three of the *telepresence schemes*: teleoperation, telecontrol and teleprogramming.

The main feature of this framework is its multitier architecture, which allows a good integration of both hardware (HW) and software (SW) elements.

Structure of the work

This work is organized as follows: In Section 2, definitions and concepts used in this work about tele-control, tele-operation and tele-programming are introduced. In Section 3, the proposed scheme based on multitier architecture is presented. The laboratory server description is given in Section 4. In Section 5, two applications of the platform are presented. The first application concerns the remote experimentation of an induction motor located in the IRCCyN laboratories in Nantes; France. The second application consists of the remote experimentation of the manipulator robot located in the CIIDIT-Mechatronic laboratories in Monterrey; Mexico. Finally, in Section 6, conclusions and recommendations are given.

2. Some concepts

Now, we introduce the concepts of teleoperation, telecontrol and teleprogramming, which will be used in the sequel.

Teleoperation is defined as the continuous, remote and direct operation of equipment (see figure 2). From the introduction of teleoperation technology, it made possible the development of interfaces capable of providing a satisfactory interaction between man and experimental equipment. On the order hand, the main aim of **telecontrol** is to extend the distance between controller devices and the equipment to the controller. Thanks to the development of the Internet, the distance between controller devices and the equipment has been increased (see figure 2).

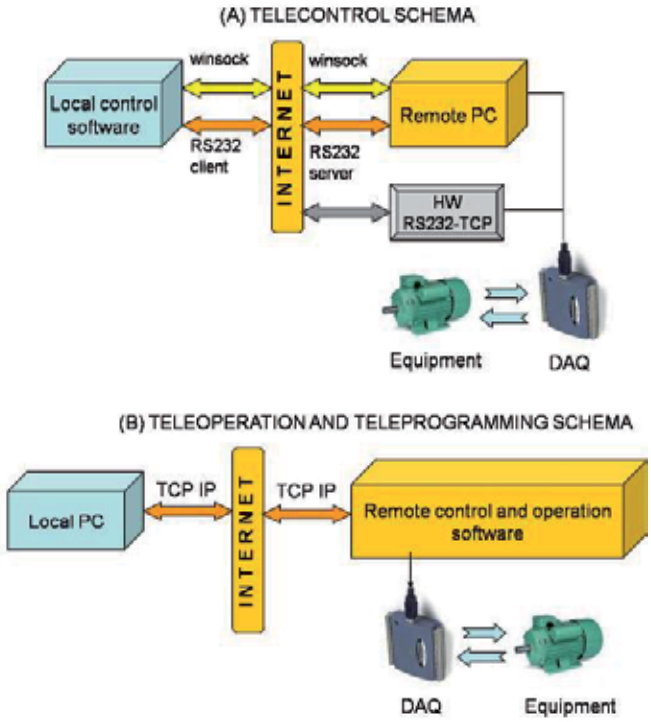


Fig. 2. Telecontrol, teleoperation and teleprogramming schema.

Figure 2.B shows a **teleoperation** scheme through the Internet working with a single channel of communication. This channel is used to change the parameters of the controller devices and/or plant. However, the effects of these changes will depend on the server layer.

Figure 2.A shows a **telecontrol** scheme through the Internet, in which the two channels of communications are required (closed-loop system), i.e. forward path *Ch1* and feedback path *Ch2*. In this case, it is necessary to maintain the stability of the closed-loop system. A solution to stability problem is that the time dalay must be less than the sampling period (Hyrun & Jong, 2005).

Furthermore, there exists a different interpretation about the **teleprogramming**. One of them is extending the distance between software programmer and the microcontroller or control board. On the other hand, it is possible to programming a remote system using two systems, called the master system and slave system, separated by the communication channel. In (Jiang et al., 2006) the teleprogramming method is based on teleoperation.

3. Framework proposed based on multitier programming

Now, we will introduce the software descriptions that are used in the proposed platform.

Figure 3 shows the tiers of the proposed framework called *Teleoption*, which has more performance than a classical telepresence framework application. *Teleoption* allows the interaction between different elements in hardware and software. Furthermore, it is possible to work under the three schemes of telepresence, i.e. teleoperation + telecontrol + teleprogramming.

The top level of the framework is the HTTP server, winsock services, webcam server and RS232 server. The second level of the framework implements the PHP script modules, DLL library and database services. All services can be shared by the VNC Server.

This distribution of software presents great advantages: i) Security in the platform, ii) several ways to transmit information from the hardware.

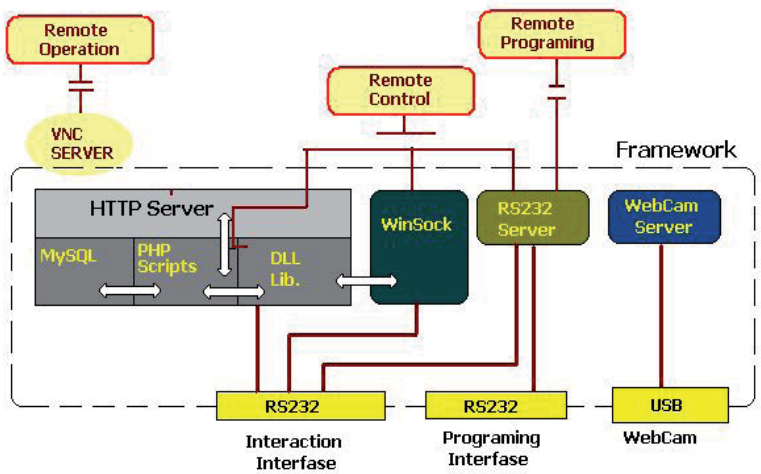


Fig. 3. Multitier architecture proposed.

Presentation tier. The HTTP Server is the presentation tier. This tier contains several Web pages with information of the platform services.

Furthermore it includes the instructions and regulation of the platform

Logic tier. In this tier, we have the programming layer. Three programming languages are used in the platform: PHP, Visual Basic and SQL. In the logic tier interacts the blocks: *i)* "PHP scripts" (which contain several programs in PHP) , *ii)* the block of the data base MySql and, *iii)* the block of the DLL libraries (designed in VBasic).

Database tier. The database tier contains information about of the platform, i.e. the users list, logbook. In fact, logic tier and database tier provide security to platform, since it is possible to use restrictions proportioned by a PHP script. This script allows the use of the platform only if the user has the permission.

Communication tier. The platform allow establish several ways of communication with the hardware: i) using *Serial Server Component* (RS232 Server), ii) using Windows sockets (Winsock) or DLL's library, and iii) using the PHP script services (see figure 4).

Serial Server Component is a software based RS232 to TCP/IP converter. RS232 Server allows any of the RS232 serial ports on the PC laboratory to interface directly to a TCP/IP network.

On the order hand, also is possible the remote access using the sockets of Windows or DLL's library. The remote user uses its own programs to send instructions to program modules of the platform.

Finally, the platform has modules designed in PHP, here, the remote user can to access to hardware using a Web page of the platform.

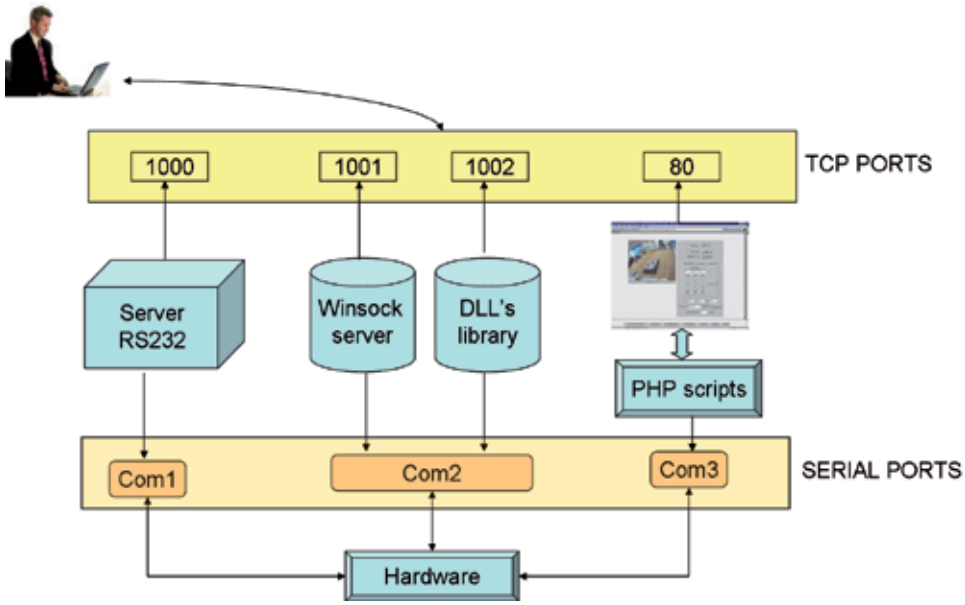


Fig. 4. Communication tier.

3.1 Operational method of the platform

When the services of **remote programming** are used, then the framework opens a communication's channel in order to share the serial services (RS232), and allows the remote programming.

If the services of **remote control** are used, then the framework opens more communication options. The first option is similar to the remote programming method, but in this case the control board and the equipment are separated, a remote communication is established by means of Internet using the services of the RS232 Server/Client.

The second alternative of remote control is the winsock option, which is similar to the last method, but the interchange of information is given by the winsock module. In this case, it is necessary to know the operation commands of the controller in order to send the information through that Internet to Winsock module, and then Winsock module will send the information to hardware.

The third option of remote control, the framework allows the access to control of the hardware using a Webpage, where the user does the work of controller. Here, the framework receives the commands of the user and sends this information to some PHP script, which sends the information to the operational layer of the multitier programming.

Finally, in the ***remote operation***, all framework are shared using the services of some VNC (Virtual Network Computer) which is a communication protocol based on RFB protocol which allows the remote access of the desktop of other computers located on the web. VNC protocol transmits the keyboard and mouse events from one computer to another, relaying the graphical screen updates back in the other direction, over a network.

4. Laboratory Server (LS) implementation

Besides the proposed framework, an architecture based on *Computers of Distributive Tasks (CDT)* is proposed. This architecture is shown in figure 5.

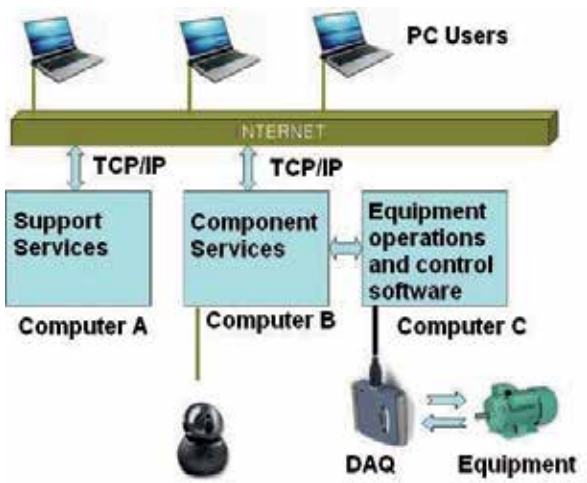


Fig. 5. Computers of Distributive Task.

Computer A allows establishing a communication both textual and oral between the local and remote user, in such way, this computer provides help on line and uses the following freeware software:

- Messenger: Textual communication and webcam.
- Skype: Oral communication, IP Telephony and videoconference.

Computer B has the task of sharing several resources through the Internet. The architecture proposed is installed in this computer. This computer uses the following software:

- Matlab/Simulink. This Software is used typically in control systems.
- ControlDesk. It is a graphical tool for controlling in real-time the equipment.
- UltraVNC server. It is software belonging to the VNC family
- LogmeIN. It is ESS software.
- TCPComm server. It is a RS232 server, which allows sharing the serial ports (COMM) of the computer. Serial port is used commonly as communication channel between PC and equipments.
- WebcamXP. Allow sharing the images from the webcams, these webcams can show the equipment details.

Computer C has an interface with the data acquisition board (DAQ), and does not share any resources on the Web. This computer is only used to share information with Computer B throughout the remote control. Furthermore, this computer protects the access to the plant (experimental equipment) in order to avoid damages caused by unauthorized users.

5. Experimental setup: Study cases

5.1 Remote experimentation of an electrical machine

The methodology described in the above section is applied to show remote access to the set-up of electrical motor located in the IRCCyN laboratory in Nantes France (figure 6), from the CIIDIT-Mechatronic laboratory in Monterrey, Mexico.

The set-up located at IRCCyN is composed of an induction motor, a synchronous motor, inverters, a real time controller board of dSPACE DS1103 and interfaces which allow to measure the position, the angular speed, the currents, the voltages and the torque between the tested machine and the synchronous motor. The motor used in the experiments has the following values: 1.5 kW normal rate power; 1430 rpm nominal angular speed; 220V nominal voltage; 7.5A nominal current; np = 2 number of pole pairs, with the motor nominal parameters: Rs = 1.633 Ohms stator resistance; Rr = 0.93 Ohms rotor resistance; Ls = 0.142H stator self-inductance; Lr = 0.076H rotor self-inductance; Msr = 0.099H mutual inductance; J = 0.0111/rad/s2 inertia (motor and load); fv = 0.0018Nm/rad/s viscous damping coefficient. The experimental sampling time T is equal to 200 s.

Furthermore, this laboratory is equipped with the remote technology described above, and can present several time delays that can appear during any real time experiments and are necessary to analyze:

- Transmission delay thought Internet (TI).
- Control algorithm computation (TC).
- Sampled time of the Data Acquisition (TS).

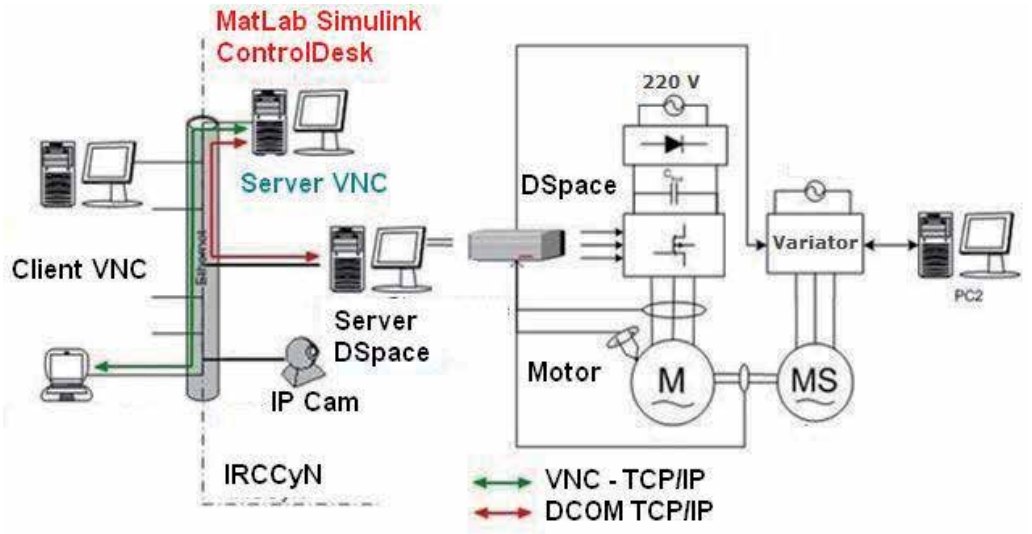


Fig. 6. IRCCyN laboratory schema.

These time delays depend on the tele-presence scheme selected. In a telecontrol scheme, the total time T = TI + TC + TS could be high and could affect the stability of the system. Nevertheless, if T = TC + TS is small, then a teleoperation scheme offers an excellent solution in remote experimentation, due to the time delay TI is not considered by the aforementioned reasons (see section 2).

Therefore, the scheme used for remote experimentation is based on teleoperation where the effects of the time delay and uncertain property is not considered in the stability of the system, because the controller and the plant are in the same layer, as shown in figure 2.

In this experiment, the time delays registered are: TI (ping) = 400 mseg. avg., TI (camera) = 3 seg. avg., TI (screen feedback, VNC) = 2 seg. avg., TC < 70 mseg.; TS = 120 seg. (DS1104).

Figure 7 shows a Mexican user, which applies a control algorithm, in order to access the remote laboratory, located in Nantes; France. From the figure 7, we can see *the computer A* showing the images sent by the webcam and the response obtained when the control algorithm is applied to the induction motor, which is transmitted by computer B using Controldesk and Matlab.

Figure 8 and 9 shows the screenshots obtained from this experiment. The first image shows the images given by webcam of the machine (with the sound), the second figure shows the Remote software ControlDesk throughout LogmeIn services.

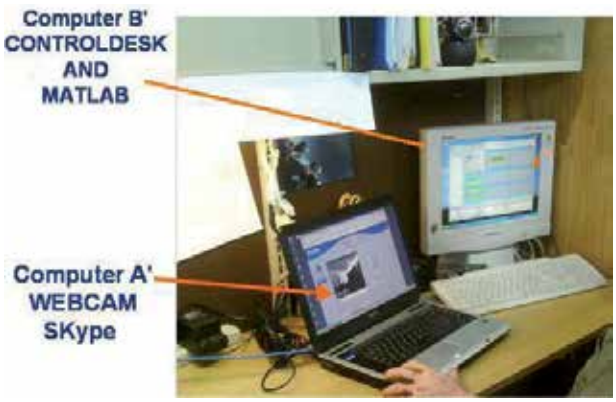


Fig. 7. Remote access by Mexican user.

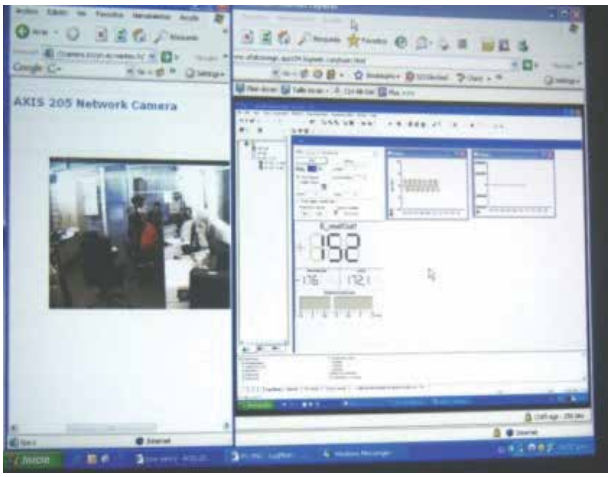


Fig. 8. Remote experimentation using LogmeIn services.

Fig. 9. Remote images of the induction motor.

5.2 Platform-setup in robotic education

It is undisputed that remote laboratories are not able to replace traditional face-to-face laboratory lessons, but they present some benefits of remote accessible experimentation:

- Flexible schedule vs. restricted schedule.
- Individual experimentation vs. group experimentation.
- Access from any computer vs. access only in the laboratory.
- Student self-learning is promoted.
- Student can use other educative means as Internet documentation, simulations, software, etc.
- The student is motivated when he is seeing his experiments and results.

This section presents another application of the architecture proposed. We emphasize that this architecture allows a remote user to access the services of control, programming and operation of robots located in the CIIDIT-Mechatronic laboratories in Monterrey; Mexico.

Teleprogramming. The objective of the teleprogramming is that the students use the BASIC microcontroller language in order to program the PICAXE microcontroller. In this platform, the student can use the basic instructions in order to program the robot: servo*, goto, serin, serout, pause, if, for*.

The student can program the PICAXE microcontroller using the flowchart method programming. Flowchart is an excellent means of pedagogy; the software shows a panoramic and graphical view of the programming sequence.

Telecontrol. The platform allows sharing the DLL resources so that the student can design programs in Visual basic, C, Matlab, or other languages. In the telecontrol option, the student can design and prove algorithms, using simulation software in local mode, subsequently if the capacity of the network is not large and it does not affect the stability of the systems, then it can be proven on-line on the robot.

Teleoperation. This platform offers the teleoperation services, so that the student can use all the services of the platform in remote mode. In this case, the platform shared the services of teleoperation using the Skype and logmeIn services.

Figure 10 showing the laboratory scheme located in CIIDIT laboratory in Mexico. The hexapod robot is acceded from the *PC Controller* Computer using two communication channels, RS232 and video. In the *PC Controller* Computer one is located the *Controller Module Server (CMS)*. The end user uses the services of the CMS in remote mode in order to control the hexapod robot.

Figure 11 showing the *screenshot* of a computed located in the IRCCyN Laboratory accessing to CIDDIT laboratory using the *LogmeIn* services.

- Figure 11.A shows the surroundings of the hexapod robot from a internal camera (eye hexapod).
- Figure 11.B presents the hexapod robot from a external camera (auxiliary camera).
- Figure 11 C shows the computers of the remote laboratory.
- Figure 11 D. showing Controller Module Client (CMC).

In the experiment, such a move-and-wait strategy is implemented of initiating control move then waiting to see the response of distant robot: then initiating a corrective move and waiting again to realize the delayed response of the distant system and the cycle repeats until the task is accomplished.

Let us define N(I) to be the number of individual moves initiated by the operator according to the move-and-wait strategy. The number N(I) depends only on the task difficulty and is independent of the delay value according to experiments (Hocayen & Spong, 2006). Consequently, the completion time, t(I), of the certain task can be calculated based on the value N(I) as follows:

$$t(I) = t_r + \sum_{i=1}^{N(I)} (t_{mi} + t_{wi}) + (t_r + t_d)N(I) + t_g + t_d \quad (1)$$

Where $t_r, t_{mi}, t_{wi}, t_g, t_d$ are human`s reaction time, movement times, waiting times after each move, grasping time and delay time introduced into communication channel, respectively.

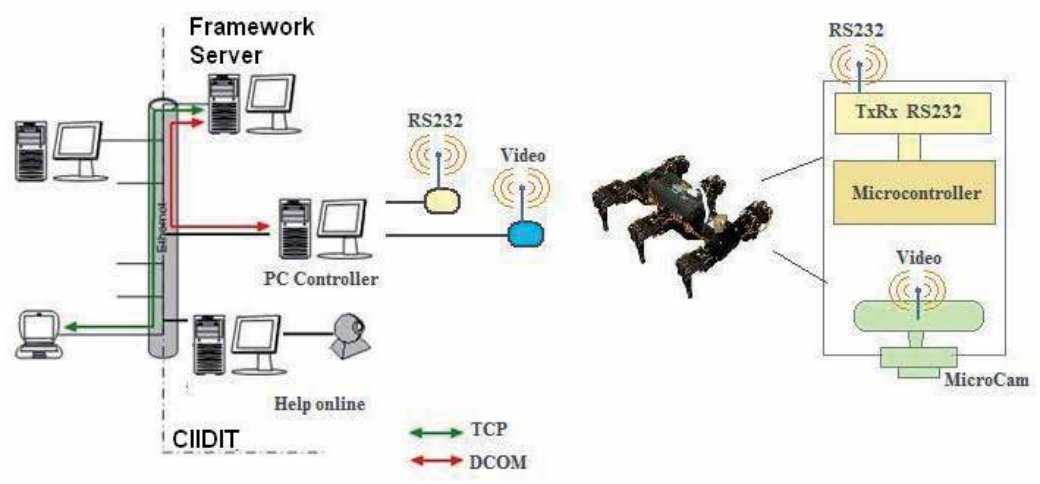


Fig. 10. CIIDIT Laboratory schema.

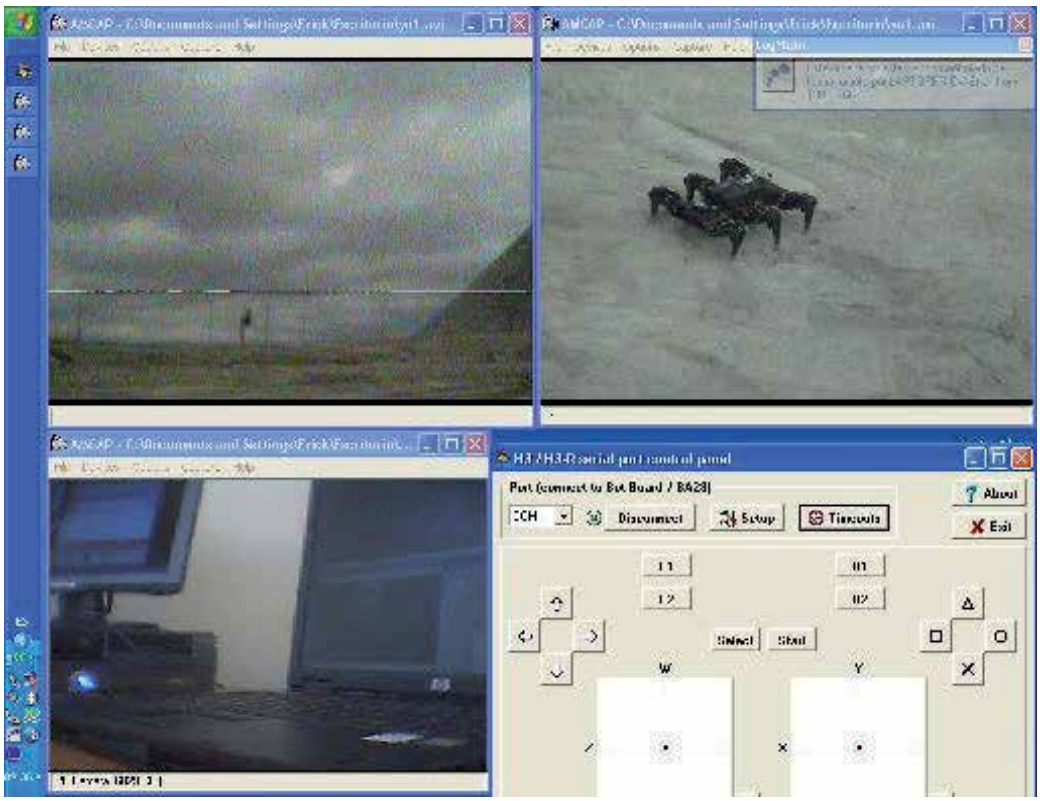


Fig. 11. Experimentation from IRCCyN, Nantes France.

6. Conclusions

In this work the capability of interfacing a large set of options with remote experimentation through the Internet has been demonstrated by the architecture based on multitier architecture.

This architecture allows the easy integration of both hardware and software, offering an excellent tool for remote experimentation, which allows the experimentation using the teleoperation, the telecontrol and teleprogramming schemes.

The main characteristic of the proposed platform has been outlined in this paper by means of a description of experiments.

7. Acknowledment

This work was supported by CONACYT, ECOS-NORD, PAICYT-UANL, Mexico and France.

8. References

Baccigalup, A.; De Capua, C.; Liccardo, A. (2006) Overview on Development of Remote Teaching Laboratories: from LabVIEW to Web Services, Instrumentation and Measurement Technology Conference, Sorrento, Italy, pp. 24-27.

Callaghan, M. J.; Harking, J.; El Gueddari, M.; McGinnity, ATM; Magure LP (2005) Client-Server Architecture for Collaborative Remote Experimentation, Procedings of the ICITA 2005, 0-7695-2316-1/05 IEEE.

Cloosterman, M.B.G.; van de Wouw, N. (2009); Heemels, W.P.M.H.; Nijmeijer, H.; Stability of Networked Control System with Uncertain Time-Variing Delay. Automatic Control, IEEE Transactions on, Volume 54, Issue 7, pp. 1575-1580.

Costas-Perez, L.; Lago, D.; Farina, J.; Rodriguez-Andina, J. (2008). Optimization o fan Industrial Sensor and Data Acquisition Laboratory Through Time Sharing and Remote Access. Industrial Electronics, IEE Transactions on, Volume 55, Issue 6, pp. 0278-0046.

Davoli, Franco; Spano, Giuseppe; Vignola, Stefano; Zappatore, Sandro. (2006) Towards Remote Laboratories With Unified Access, IEEE Transactions on Instrumentation and Measurement", Vol 55, No. 5.

Gomez, Luís; Garcia, Javier (2007); Advances on Remote Laboratories and e-learning experencies. Deusto Publicaciones, ISSB 975-84-9830-662-0

Hokayen, Peter F.; Spong, Mark W. (2006) Bilateral teleoperation: An historical survey, Automatica 42 : 2035-2057

Huijun Gao; Tomgwen Chen; James Lam (2008); A new delay system approach to network-based control. Automatica, Volume 44; Issie 1, pp. 39-52

Hyun, Chul Cho; Jong, Hyeon Parck (2005) Stable bilateral teleoperation under time delay using robust impedance control. Mechatronic, Vol. 15: 611-625.

Jiang, Zainan; Xie, Zong; Wang, Bin; Wang, Jie; Liu, Hong (2006) A teleprogramming Methos for Interned-based Teleoperation. International Conference on Robotics and Biomimetics, Dec. 17-20, Kuynming China.

Rapuano, Sergio; Zoino, Francesco (2005) A learning Management System Including Laboratory Experiments on Measurement Instrumentation, IMTC 2005, Instrumentation and Measurement Technology Conference, Ottawa, Canada, pp. 17 - 19 .

Restivo, M.T.; Mendes, J.; Lopes, A.M.; Silva, C.M.; Chouzal, F (2009). A Remote Laboratory in Enginnering Measurement. Industrial Electronics, IEEE Transactions on. Volume 56, Issue 12, pp. 4836-4843.

Wang, Meng; James N.K (2005) Interactive Control for Internet-based Mobile Robot Teleoperation, Robotics and Autonomous System 52, pp. 160-179.

Wu, Y. L; Chan, T.; Jong B.S.; Lin, T.W. (2008) A Web-based virtual reality physic laboratory", In Pro 3rd IEEE ICALT, Athenas Grerce, pp.455.

11

Multicriteria Optimization in Telecommunication Networks Planning, Designing and Controlling

Valery Bezruk, Alexander Bukhanko,
Dariya Chebotaryova and Vacheslav Varich
Kharkov National University of Radio Electronics
Ukraine

1. Introduction

Modern telecommunication networks, irrespectively of their organization and type of the transmitted information, become more complex and possess many specific characteristics. The new generation of telecommunication networks and systems support a wide range of various communication-intensive real-time and non real-time various applications. All these net applications have their own different quality-of-service requirements in terms of throughput, reliability, and bounds on end-to-end delay, jitter, and packet-loss ratio etc. Thus, telecommunication network is a type of the information system considered as an ordered set of elements, relations and their properties. Their unique setting defines the goal searching system.

For such a type of information system as a telecommunication network it is necessary to perform a preliminary long-term planning (with structure designing and system relation defining) and a short-term operating control within networks functioning. The problem of the optimal planning, designing and controlling in the telecommunication networks involves: definition of an initial set of decisions, formation of a subset of system permissible variants, definition of an optimal criteria, and also a choice of the structure variants and network parameters, optimal by such a criteria. It is the task of a general decision making theory reduced to the implementation of some choice function of the best (optimal) system based on the set of valid variants. For the decision making tasks the following optimizing methods can be used: scalar and vector optimization, linear and nonlinear optimization, parametric and structure optimization, etc (Figueira, 2005; Taha, 1997; Saaty, 2005). We propose a method of the multicriteria optimization for optimum variants choice taking into account the set of quality indicators both in long-term and short-term planning and controlling.

The initial set of permissible variants of a telecommunication network is being formed through the definition of the different network topologies, transmission capacities of communication channels, various disciplines of service requests applied to different routing ways, etc. Obtained variants of the telecommunication network construction are estimated

on a totality of given metrics describing the messages transmission quality. Thus, the formed set of the permissible design decisions is represented in the space of criteria ratings of quality indicators where, used of unconditional criteria of a preference, the subset of effective (Pareto-optimal) variants of the telecommunication network is selected. On a final stage of optimization any obtained effective variants of the network can be selected for usage. The unique variant choice of a telecommunication network with introducing some conventional criteria of preference as some scalar goal function is also possible.

In the present work some generalizations are made and all stages of solving multicriteria problems are analyzed with reference to telecommunication networks including the statement of a problem, finding the Pareto-optimal systems and selecting the only system variant. This chapter also considers the application particularities of multicriteria optimization methods at the operating control within telecommunication systems. The investigation results are provided on the example of solving of a particular management problem considering planning of cellular networks, optimal routing and choice of the speech codec, controlling network resources, etc.

2. Theoretical investigation in Pareto optimization

As far as the most general case is concerned, the system can be thought of as an ordered set of elements, relationships and their properties. The uniqueness of their assignment serves to define the system fully, notably, its structure and efficiency. The major objective of designing is to specify and define all the above-listed categories. The solution of this problem involves determining an initial set of solutions, generating a subset of pemissible solutions, assigning the criteria of the system optimality and selecting the system, which is optimal in terms of a criteria.

2.1 The problem statement in optimization system

It is assumed that the system $\varphi = (s, \vec{\beta}) \in \Phi_D$ is defined by the structure s (a set of elements and connections) and by the vector of parameters $\vec{\beta}$. A set of input actions X and output results Y should be assigned for an information system. This procedure defines the system as the mapping $\varphi : X \to Y$. The abstract determination of the system in the process of designing is considered to be exact. In particular, when formalizing the problem statement, a mathematical descripton of the working conditions (of signals, interferences) and of the functional purpose of a system (solutions obtained at the system output) are to be given, which, in fact, determine the variant of the system $\varphi \in \Phi$.

In particular, the limitations given on conditions of work, on the structure $s \in S_D$ and parameters $\beta \in B_D$, as well as on values of the system quality indicators define the subset of permissible project solutions $\Phi_a = S_a \times B_a$. Diverse ways of assigning a set of allowable are possible, in particular:

- implicit assignment using the limitations upon the operating conditions formulated in a rigorous mathematical form;
- enumeration of permissible variants of the system;
- determination of the formal mechanism for generating the system variants.

The choice of the optimal criteria is related to the formalization of the knowledge about an optimality. There exist two ways of describing the customer's preference of one variant to the other, i.e. ordinal and cardinal .

An ordinal approach is order-oriented (better-worse) and is based on introducing certain binary relations on a set of permissible alternatives. In this case the customer's preference is the binary relation R on the set Φ_D which reflects the customer's knowledge that the alternative φ' is better than the alternative: φ'' : $\varphi' R \varphi''$.

Assume that a customer sticks to a certain rigorous preference $\succ$, which is asymmetric and transitive, as he decides on a set of permissible alternative Φ_D. The solution $\varphi_0 \in \Phi_D$ is called optimal with respect to $\succ$, unless there are other solution $\varphi \in \Phi_D$ for which $\varphi \succ \varphi^{(0)}$ holds true. A set of all optimal solutions in relation to $\succ$ is denoted by $\mathrm{opt}_{\succ}\Phi_D$. A set of optimal solutions can comprise the only element, a finite or infinite number of elements as a function of the structure of a permissible set or properties of the relation $\succ$. If the discernibility relation coincides with that of equality =, then the set $\mathrm{opt}_{\succ}\Phi_D$ (provided it is not empty) contains the only element.

A cardinal approach to describe the customer's preference assigns to each alternative $\varphi \in \Phi_D$, a certain number U being interpreted as the utility of the alternative φ. Each utility function determines a corresponding order (or a preference) R on die set $\Phi_D (\varphi' R \varphi')$ if and only if $U(\varphi') \geq U(\varphi'')$. In this case they say that the utility function $U(\cdot)$ is a preference indicator R. In point of fact this approach is related to assigning a certain scalar-objective function (a conventional preference criteria) whose optimization in a general case may result in the selection of the only optimal variant of the system.

The choice of the optimal criteria is based on formalizing the knowledge of a die system customer (i.e. a person who makes a decision) about its optimality. However, one often fails to formalize the knowledge of a decision-making person about the system optimality rigorously. Therefore, it appears impossible to assign the implicitly of the scalar optimal criteria resulting in the choice of the only decision variant $\varphi^{(0)} = \underset{\varphi \in \Phi_D}{\mathrm{extr}} [U(\varphi)]$, where $U(\varphi)$ is a certain objective function of the system utility (or usefulness). Therefore, at the initial design stages the system is characterized by a set of objective functions:

$$\vec{k}(\varphi) = (k_1(\varphi), ..., k_i(\varphi), ..., k_m(\varphi)), \tag{1}$$

which determines the influence of the structure s and the parameters $\vec{\beta}$ of the variant of the system $\varphi = (s, \vec{\beta})$ upon the system quality indicators. In this connection one has to deal with the newly emerged issues of optimizing approaches in terms of a collection of quality indicators, which likewise are called the problems of multicriteria or vector optimization. Basically, the statement and the solution of a multicriteria problems is related to replacing (approximation) customer's knowledge about the system optimality with a different optimality conception which can be formalized as a certain vector optimal criteria (1) and, consequently, the problem will be solved through the effective optimization procedure.

2.2 Forming a set of permissible variants of a system

When optimizing the information systems, as their decomposition into subsystems can be assigned, it would be judicious to proceed from the morphological approach which is widely applied in designing complicated systems. In this context it is assumed that any variant of a system has a definite structure, i.e. it consists of the finite number of elements (subsystems), and the distribution of system functions amongst them can be performed by the finite number of methods.

Now consider the peculiar features of generating the structural set of permissible variants of a system. Let us assume that the functional decomposition of the system into a set of elements is

$$\{\varphi_j,\ j=\overline{1,L},\ \bigcup_{j=1}^{L}\varphi_j=\varphi\}.$$

What is considered to be assigned is as follows: a finite set of elements of the system E as well as the splitting of the set E into L morphological classes $\sigma(l), l=\overline{1,L}$ such as $\sigma(l)\cap\sigma(l')=\varnothing$ at $l\neq l'$.

A concept of the morphological space $\Lambda\subseteq 2^{\varepsilon}$ is introduced, its elements being the morphological variant of the system $\varphi=(\varphi_1,\varphi_2,\ldots,\varphi_L)$. Each morphological variant φ is a certain set of representatives of the classes $\varphi(l)\in\sigma(l)$. Here for all $\varphi\in\Lambda$ and for any $l=\overline{1,L}$ the set $\varphi\in\Lambda$ contains a single element.

Under the assumption that there exist a multitude of alternative model of implementing each subsystem φ_{lk}, $k=\overline{1,L}$, $l=\overline{1,L}$, the following morphological table can be specified:

Morphological classes	Possible models of implementing the system elements	Number of modes of implementing the system
$\sigma(1)$	$\varphi_{11}[\varphi_{12}]\varphi_{13}\ldots\varphi_{1K_1}$	K_1
$\sigma(2)$	$\varphi_{21}\varphi_{22}\varphi_{13}\ldots[\varphi_{2K_2}]$	K_2
……	……	……
$\sigma(l)$	$\varphi_{l1}\varphi_{l2}[\varphi_{l3}]\ldots\varphi_{lK_l}$	K_l
……	……	……
$\sigma(L)$	$[\varphi_{L1}]\varphi_{L2}\varphi_{L3}\ldots\varphi_{LK_L}$	K_L

Table 1. Morphological table.

As an example (see table 1), a q-th morphological variant of the system $\varphi^q=\langle\varphi_{12},\varphi_{2K_2},\ldots,\varphi_{l3},\ldots,\varphi_{L1}\rangle$ that determines the system structure is distinguished. The total number of all possible morphological variants of the system is generally determined as $Q=\prod_{l=1}^{L}K_l$.

When generating a set of permissible variants Φ_D one has to allow for the constraints upon the structure, parameters and technical realization of elements and the system as a whole as well as for the permissible combination of elements connections and constraints up on the value of the quality indicators of the system as a whole.

Here, there exist conflicting requirements. On the one hand, it is desirable to present all conceivable variants of the system in their entirety so as not to leave out the potentially best variants. On the other hand, there are limitations specified by the permissible expenditures (of time and funds) on the designing of a system.

After a set of permissible variant of a system has been determined in terms of a particular structure, the value of the quality indicators is estimated, a set of Pareto-optimal variants is distinguished and gets narrowed down to the most preferable one.

2.3 Finding the system Pareto-optimal variants

As a collection of objective functions is being introduced, each variant of the system φ is mapped from a set of permissible variants Φ_D into the criteria space of estimates $V \in R^m$:

$$V = \vec{K}(\Phi_D) = \{\vec{v} \in R^m \mid \vec{v} = \vec{k}(\varphi), \varphi \in \Phi_D\}. \quad (2)$$

In this case to each approach φ corresponds its particular estimate of the selected quality indicators $\vec{v} = \vec{k}(\varphi)$ (2) and, vice versa, to each estimate corresponds an approach (in a general way, a single approach is not obligatory).

To the relation of the rigorous preference $\succ$ on the set Φ_D corresponds the relation $\succ$ in the criteria space of estimates V. According to the Pareto axiom, for any two estimates $\vec{v}', \vec{v}'' \in V$ satisfying the vector inequality $\vec{v}' \geq \vec{v}''$, the relation $\vec{v}' \succ \vec{v}''$ is always obeyed. Besides, according to the second Pareto action for any two approaches $\varphi', \varphi'' \in \Phi_D$, for which $\vec{k}(\varphi') \geq \vec{k}(\varphi'')$ is true, the relation $\varphi' \succ \varphi''$ always occurs. The Pareto axiom imposes definite limitations upon the character of the preference in multicriteria problem.

It is desirable for a customer to obtain the best possible value for each criteria. Yet in practice this case can be rarely found. Here, it should be emphasized that the quality indicators (objective function) of the system (1) may be of 3 types: neutral, consistent with one another and competing between one other. In the first two instances the system optimization can be performed separately in terms of each of indicators. In the third instance it appears impossible to arrive at a potential value of each of the individual indicators. In this case one can only attain the consistent optimum of introduced objective functions – the optimum according to the Pareto criteria which implies that each of the indicators can be further improved solly by lowering the remaining quality indicators of the system. To the Pareto optimum in the criteria space corresponds a set of Pareto-optimal estimates that satisfy the following expression:

$$P(V) = \mathrm{opt}_{\geq} V = \{\vec{k}(\varphi^0) \in R^m \mid \forall \vec{k}(\varphi) \in V : \vec{k}(\varphi) \geq \vec{k}(\varphi^0)\}. \quad (3)$$

An optimum based on the Pareto criteria can be found either directly according to (3) by the exhaustive search of all permissible variants of the system Φ_D or with the use of special procedures such as the weighting method, methods of operating characteristics.

With the Pareto *weighting method* being employed. The optimal decisions are found by optimizing the weighted sum of objective functions

$$\underset{\varphi\in\Phi_D}{extr}[k_p(\varphi)=\lambda_1 k_1(\varphi)+\lambda_2 k_2(\varphi)+...+\lambda_m k_m(\varphi)], \tag{4}$$

in which the weighting coefficients $\lambda_1,\lambda_2,\ldots,\lambda_m$ are selected from the condition $\lambda_i>0, \sum_{i=1}^{m}\lambda_i=1$. The Pareto-optimal decisions are the system variants that satisfy eq. (4) with different permissible combination of the weighting coefficients $\lambda_1,\lambda_2,\ldots,\lambda_m$. When solving this problem one can observe the variation in the alternative systems $\varphi=(s,\vec{\beta})\in\Phi_D$ within the limits of specified.

The method of operating characteristics consists all the objective functions, except for a single one, say, the first one, are transferred into a category of limitations of an inequality type, and its optimum is sought on a set of permissible alternatives

$$\underset{\varphi\in\Phi_D}{extr}[k_1(\varphi)],\ k_2(\varphi)=K_{2\varphi};\ k_3(\varphi)=K_{3\varphi},...,\ k_m(\varphi)=K_{m\varphi}. \tag{5}$$

Here $K_{2\varphi},K_{3\varphi},...,K_{m\varphi}$ are the certain fixed, but arbitrary quality indicators values.

The optimization problem (5) is solved sequentially for all permissible combinations of the values $K_{2\varphi}\le K_{2D}, K_{3\varphi}\le K_{3D},..., K_{m\varphi}\le K_{mD}$. In each instance an optimal value of the indicator k_{1opt} is sought by variations $\varphi\in\Phi_D$. As a result a certain multidimensional working space in the criteria space is sought

$$k_{1opt}=f_p(K_{2\varphi},K_{3\varphi},...,K_{m\varphi}). \tag{6}$$

If the found relation (6) is monotonously decreasing in nature for each of the arguments, the working surface coincides with a Pareto-optimal surface. This surface can be connected, nonconnected and just a set of isolated points.

It should be pointed out that each point of the pareto-optimal surface offers the property of a m-fold optimum, i.e. this point checks with a potentially attainable (with variation $\varphi\in\Phi_D$) value of one of the indicators k_{1opt} at the fixed (corresponding to this point) value of other $(m-1)$ quality indicators. The Pareto-optimal surface can be described by any of the following relationships

$$k_{1opt}=f_{no}^{1}(k_2,k_3,...,k_m),...,k_{mopt}=f_{no}^{m}(k_1,k_2,...,k_{m-1}), \tag{7}$$

which represent the multidimensional diagram of the exchange between the quality indicators showing the way in which the potentially attainable value of the corresponding indicator depends upon the values of other indicators.

Thus, the Pareto-optimal surface connects the potentially attainable values of index is Pareto-optimum consistent, generally dependent and competing quality indicators Therefore, with

the Pareto-optimal surface in the criteria space being obtained, the multidimensional potential characteristics of the system and related multidimensional exchange diagram are found.

It should be noted that they are different types of optimization problems depending upon the problem statement.

Discrete selection. The initial set Φ_D is specified by a finite number of variants of constructing the system $\{\varphi_l, l=\overline{1,L_D}, \varphi \in \Phi_D\}$. It is required that set of Pareto-optimal variants of the system $\text{opt}_{\succ}\Phi_D$. should be selected.

Parametric optimization. The structure of the system S_D is specified. It is necessary to find the magnitude of the vectors $\vec{\beta}^0 \in B_D$ at which $\varphi = (s,\vec{\beta}) \in \text{opt}_{\succ}\Phi_D$.

Structural-parametric optimization. It is necessary to synthesize the structure $s \in S_D$ and to find the magnitude of the vector of the parameters $\vec{\beta} \in B_D$ at which $\varphi = (s,\vec{\beta}) \in \text{opt}_{\succ}\Phi_D$.

The first two types of problems have been adequately developed in the theoiy of multicriteria optimization. The solution of the third-type problems is most complicated. To synthesize the Pareto-optimal structure and find the optimal parameters a set of functionals $k_1(s,\vec{\beta}), k_2(s,\vec{\beta}),..., k_m(s,\vec{\beta})$ is to be optimized. Yet optimizing functionals even in a scalar case appears to be a rather challenging task from both the mathematical and some no less importants standpoints. In the case of a vector the solution to these types of problems becomes still more complicated. Therefore, in designing the systems with regard to a set of the quality indicators one has to simplify the optimization problem by decomposing the system into simpler subsystems, to reduce the number of quality indicators as the system structure is being synthesized.

If the set of Pareto-optimal systems variants, which has been found following the optimization procedure, turned out to be a narrow one, then any of them can be made use of as an optimal one. In this case the rigorous preference relation $\succ$ may be thought of as coinciding with the relation $\geq$ and, therefore, $\text{opt}_{\succ}V = P(V)$.

However, in practice the set $P(V)$ proves to be sufficiently wide. This implies that the relations $\succ$ and $\geq$ (although they are connected through the Pareto axiom) do not show a close agreement. Here, the inclusions $\text{opt}_{\succ}V \subset P(V)$ and $\text{opt}_{\succ}\Phi_D \subset P_{\tilde{k}}(\Phi_D)$ are valid. Therefore, we will have to deal with an emerging problem of narrowing the found Pareto-optimal solutions involving additional information about the relation of the customer's rigorous preference. Yet the ultimate selection of optimal approaches should only be made within the limits of the found set of Pareto-optimal solution.

2.4 Narrowing of the set of Pareto-optimal solutions down to the only variant of a system

The formal model of the Pareto optimization problem does not contain any information to select the only alternative. In this particular instance a set of permissible variants gets narrowed only to a set of Pareto-optimal solution by eliminating the worse variants with respect to a precise variant.

However, the only variant of a system is normally to be chosen to ensure the subsequent designing stages. It is just for this reason why one feels it necessary to narrow the set of Pareto-optimal solutions down to the only variant of a system and to make use of some additional information about a customer's preference. This type of information is produced following the comprehensive analysis of Pareto-optimal variants of a system, particularly, of a structure, parameters, operating characteristics of the obtained variants of a system, a relative importance of input quality indicators, etc. Some additional information thus obtained concerning the customer's preferences is employed to construct choice function (an objective scalar function) whose optimization tends to select the sole variants of a system.

In order to solve the problem of narrowing a set of Pareto-optimal solution a diversity of approaches, especially those based on the theory of utility, the theory of fuzzy sets, etc. Now let us take a brief look at some of them.

The selection of optimal approaches using the scalar value function. One of the commonly used methods of narrowing a set of Pareto-optimal solution is constructing the scalar value function, which, if applied, gives rise to selecting one of the optimal variants of a system.

The numerical function $F(v_1, v_2, ..., v_m)$ of m variables is referred to as the value (utility) function for the relation $\succ$ if for the arbitrary estimates $\vec{v}', \vec{v}'' \in V$ the inequality $F(\vec{v}') > F(\vec{v}'')$ occurs if and only if $\vec{v}' \succ \vec{v}''$. If there exists the function of utility $F(\vec{v})$ for the relation $\succ$, then it is obvious that

$$\mathrm{opt}_{\succ} V = \{\vec{v}^0 \in V : F(\vec{v}^0) = \max_{\vec{v} \in V} F(\vec{v})\}$$

and finding an optimal estimate boils down to solving the single-criteria problem of optimizing the function $F(\vec{v})$ on the set V. The value function of the type

$$F(v_1, v_2, ..., v_m) = \sum_{j=1}^{m} c_j f_j(v_j), \tag{8}$$

where c_j is the scaling factor, $f_j(v_j)$ are the certain unidimensional value function which are the estimates of usefulnen of the system variant φ in terms of the index $k_j(\varphi)$.

The construction of the value function (8) consists in estimating the scale factors, forming unidimensional utility function $f_j(v_j)$ as well as in validating their independence and consistency. Here, use is made of the data obtained from interrogating a customer. Special interrogation procedures and program packages intended to acquire some additional information about the customer's preferences have been worked out.

The selection of optimal approaches based upon the theory of fuzzy sets. This procedure is based on the fact that due to the apriori uncertainty with regard to the customer's preference, the concept such as "the best variant of a system" cannot be accurately defined. This concept may be thought of as constituting a fuzzy set and in order to make an estimate of the system, the basic postulates of the fuzzy- set theory can be employed.

Let X be a certain set of possible magnitudes of a particular quality indicator of a system. The fuzzy set G on the set X is assigned by the membership function $\xi_G : X \to [0,1]$ which brings the real number ξ_G over the interval [0,1] in line with each element of the set X. The value ξ_G defined the degree of membership of the set X elements to the fuzzy set G. The nearer is the value $\xi_G(x)$ to unity, the higher is the membership degree. The membership function $\xi_G(x)$ is the generalization of the characteristic function of sets, which takes two values only : 1 – at $x \in G$; 0 – at $x \notin G$. For discrete sets X the fuzzy set G is written as the set of pairs $G = \{x, \xi_G(x)\}$.

Thus, according to the theory of fuzzy sets each of the quality indicators can be assigned in the form of a fuzzy set

$$k_j = \{k_j, \xi_{k_j}(k_j)\},$$

where $\xi_{k_j}(\circ)$ is the membership function of the specific value of the j-th index to the optimal magnitude.

This type of writing is highly informative, since it gives an insight into its physical meaning and "worth" in relation to the optimal (extreme) value which is characterized by the membership function $\xi_{k_j}(\circ)$.

The main difficulty over the practical implementation of the considered approach consists in choosing the type of a membership function. In some sense the universal form of the membership function being interpreted in terms of the theory of fuzzy sets with regard to the collection of indicators is written as:

$$\xi_{\bar{k}}(k_1, k_2, ..., k_m) = \frac{1}{m}\left\{\sum_{l=1}^{m}[\xi_{k_j}(k_j)]^{\beta}\right\}^{\frac{1}{\beta}}. \quad (9)$$

The advantage of this form is that depending upon the parameter β a wide class of functions is implemented. These functions range from the linear additive form at $\beta = 1$ to the particularly nonlinear relationships at $\beta \to \infty$.

It should be pointed out that with this particular approach it is essential that the information obtained from a customer by an expert estimates method be used to pick out a membership function and a variety of coefficients.

Selecting optimal approaches at quality indicators strictly ordered in terms of the level of their importance. Occasionally it appears desirable for a customer to obtain die maximum magnitude of one of the indicators, say, k_1 even at the expense of the "lasses" for the remaining indicators. This means that the indicator k_1 is found to be more important than other indicators.

In addition, there may be the case where the whole set of indicators $k_1, k_2, \ldots, k_m$ is strictly ordered in terms of their importance such k_1 is more important that other indicators $k_1, k_2, \ldots, k_m$; k_2 is more essential than all the indicators $k_1, k_2, \ldots, k_m$, etc. This corresponds the instance where the lexico-graphical relation lex is employed when a comparison is made between the estimates of approaches. Now we give the definition of the above relation.

Let there be two vectors of estimates $\vec{v}, \vec{v}' \in V \subset R^m$. The lexico-graphical relation $\overset{lex}{\succ}$ is determined in the following way: the relation $\overset{lex}{\succ}$ occurs if and only if one the following conditions is satisfied.

$$1)\ v_1' > v_1'';$$
$$2)\ v_1' = v_1''; v_2' > v_2''$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$m)\ v_j' = v_j'', j = 1,2,\ldots,m-1, v_m' > v_m'',$$
$$v' = (v_1', v_2', \ldots, v_m');\ v^n = (v_1^n, v_2^n, \ldots, v_m^n).$$

In this case the components $v_1, v_2, \ldots, v_m$, i.e. the estimates of the system quality indicators $k_1(\varphi), k_2(\varphi), \ldots, k_m(\varphi)$ are said to be strictly order in terms of their importance. As the relation $\vec{v}' lex \vec{v}''$ is satisfied they say that from the lexico-graphical stand point the vector $\vec{v}'$ is greater than the vector $\vec{v}''$. At $m-1$ the lexico-graphical relation coincides with the relation $\succ$ on the subset of real numbers.

In determining the lexico-graphical relation a major role is played by the order of enumerating quality indicators. The change in the numeration of quality indicators give rise to a different lexico-graphical relation.

3. Practical usage

Let us consider some practical peculiarities of an application of multicriteria optimization methods within a long-term and short-term planning, designing and controlling. In the examined examples of telecommunication networks operation and estimation of the quality indicators values is probed on mathematical models implemented on a computer using the packets of specific simulation modeling.

3.1 Telecommunication network variant choice

In particular, we considered features of an application of multicriteria optimization methods on the example of the packet switching network. For such a task the mathematical model of full-connected topology of a network was implemented. There was performed the simulation modeling of different variants of data transmission in the indicated network and the quality indicators estimates for each variant were obtained (Bezruk et al., 2008).

Pareto-optimal variants of the network were obtained with the methods of vector optimization and, among them, there was selected the single optimal variant of the network (fig. 1). The results of the optimization were used for the task of the network control when framing optimal control actions.

Thus, the control device collects the information on the current condition of the network and develops Pareto-optimal control actions which are directed to a variation of mechanisms of the arrival requests service and paths of packet transmission through the network.

The structure of the model, realized with a computer, includes simulators of the messages with a Poisson distribution and given intensities, procedures of the messages packing, their

transmission through the communication channels. The procedures of the messages packing have simulated a batch data transmission with a mode of the window load control.

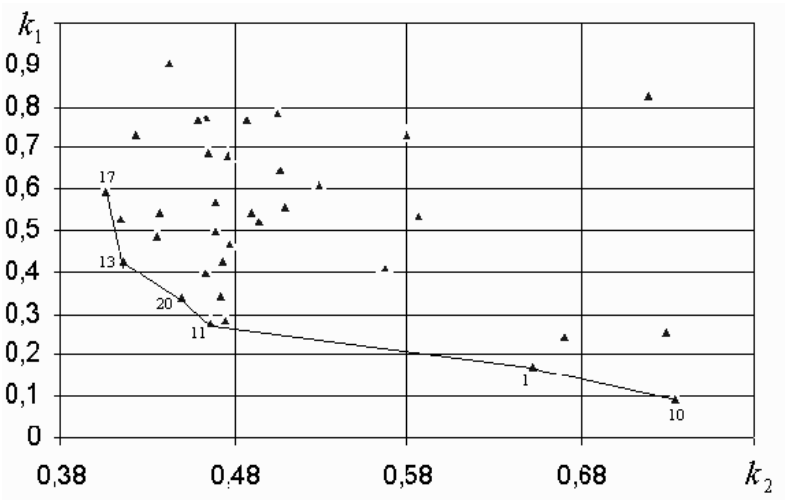


Fig. 1. Choice of Pareto-optimal variants of the telecommunication network.

The procedures of a packet transmission were simulated by the processes of transfer using duplex communication channels with errors. The simulation analysis of the transfer delays was stipulated at a packet transmission in the communication lines connected with final velocity of signals propagation in communication channels, fixed transmission channel capacity and packets arrival time in the queue for their transfer trough the communication channel.

Different variants of the telecommunication network functioning were realized at the simulation analysis, they differed in disciplines of service in the queues, ways of routing in a packet transmission and size of the window of the transport junction. In the considered example thirty six variants of the network functioning were obtained. Network functioning variants were estimated by the following quality indicators: average time of deliveries $k_1 = \overline{T}$ and average probability of message loss $k_2 = \overline{P}$. These quality indicators had contradictory character of interconnection. The obtained permissible set of network variants is presented in a criteria space (fig. 1). The subset of the Pareto-optimal network operation is selected by the exclusion of the inferior variants. The left low bound set of the valid variants corresponds to Pareto-optimal variants. Among Pareto-optimal variants of the network Φ_0 was selected a single variant from the condition of a minimum of the introduced resulting quality indicator $k_{pn} = C_1 k_1 + C_2 k_2$. For the case $C_1 = 0,4$, $C_2 = 0,6$ the single variant 11 was selected; the discipline service of the requests (in the random order) was established for it as well as the way of routing (weight method) and size of the "transmission window" (equal 8).

The given task is urgent for practical applications being critical to the delivery time (in telecommunication systems of video and voice intelligences, systems of the banking terminals, alarm installations, etc).

3.2 Multicriteria optimization in radio communication networks designing

Let us consider some practical aspects of multicriteria optimization methods when planning radio communication networks, on an example of cellular communication network (CCN). The process of finding CCN optimal variants includes such stages:

- setting the initial set of the system variants differed in the following terms: radio standards, the engaged frequency band, the number and activity of subscribers, covered territory, sectoring and the height of antennas, the power of base station transmitters, the parameter of radio wave attenuation, etc;
- separation of the permissible set of variants with regard of limitations on the network structure and parameters, limitation on the value of the quality indicators;
- choice of the subset of Pareto-optimal CCN variants;
- analysis of obtained Pareto-optimal CCN variants;
- choice of a single CCN variant.

In the considered example there was formed a set of permissible variants of CCN (GSM standard), which were defined by different initial data including the following ones: the planned number of subscribers in the network; dimensions of the covered territory (an area); the activity of subscribers at HML (hour of maximum load); the frequency bandwidth authorized for the network organization; sizes of clusters; the permissible probability of call blocking and percentage of the time of the communication quality deterioration.

The following technical parameters of CCN were calculated by a special technique.

1. The general number of frequency channels authorized for deployment of CCN in the given town, is defined as

$$N_k = \mathrm{int}(\triangle F / F_k),$$

where F_k is the frequency band.

2. The number of radio frequencies needed for service of subscribers in one sector of each cell, is defined as

$$n_s = \mathrm{int}(N_k / C \cdot M).$$

3. A value of the permissible telephone load in one sector of one cell or in a cell (for base stations incorporating antennas with the circular pattern) is defined by the following relationships

$$A = n_O \left[1 - \sqrt{1 - \left(P_{sl} \sqrt{\pi n_O / 2} \right)^{1/n_0}} \right] \text{ at } P_{sl} \le \sqrt{\frac{2}{\pi n_o}};$$

$$A = n_O + \sqrt{\frac{\pi}{2} + 2\pi_O \ln\left(P_{sl} \sqrt{\pi n_O / 2} \right)} - \sqrt{\frac{\pi}{2}} \text{ at } P_{sl} > \sqrt{\frac{2}{\pi n_o}},$$

where $n_0 = n_s \cdot n_a$; n_a is the number of subscribers which can use one frequency channel simultaneously. The value is defined by standard.

4. The number of subscribers under service of the base station, depending on the number of sectors, permissible telephone load and activity of subscribers

$$N_{aBTS} = M\,\mathrm{int}(A/\beta).$$

5. The necessary number of the base stations at the given territory of covering, is defined as

$$N_{BTS} = \mathrm{int}(N_a / N_{aBTS}).$$

where N_a is the given number of subscribers to be under service of the cellular communication network.

6. The cell radius, under condition that the load is uniformly distributed over the entire zone, is defined by the formula

$$R = \sqrt{\frac{1,21 \cdot S_0}{\pi N_{BTS}}}.$$

7. The value of the protective distance between BTS with equal frequency channels, is defined as

$$D = R\sqrt{3C},$$

and other parameters such as the necessary power at the receiver input, the probability of error in the process of communication session, the efficiency of radio spectrum use, etc.

Finding the subset of Pareto-optimal network variants is performed in criteria space of the quality indicators estimates. A single variant of CCN was chosen with the use of the conditional criteria of preference by finding the extreme of the scalar criteria function as $c_i = \frac{1}{7},\ i = \overline{1,7}$.

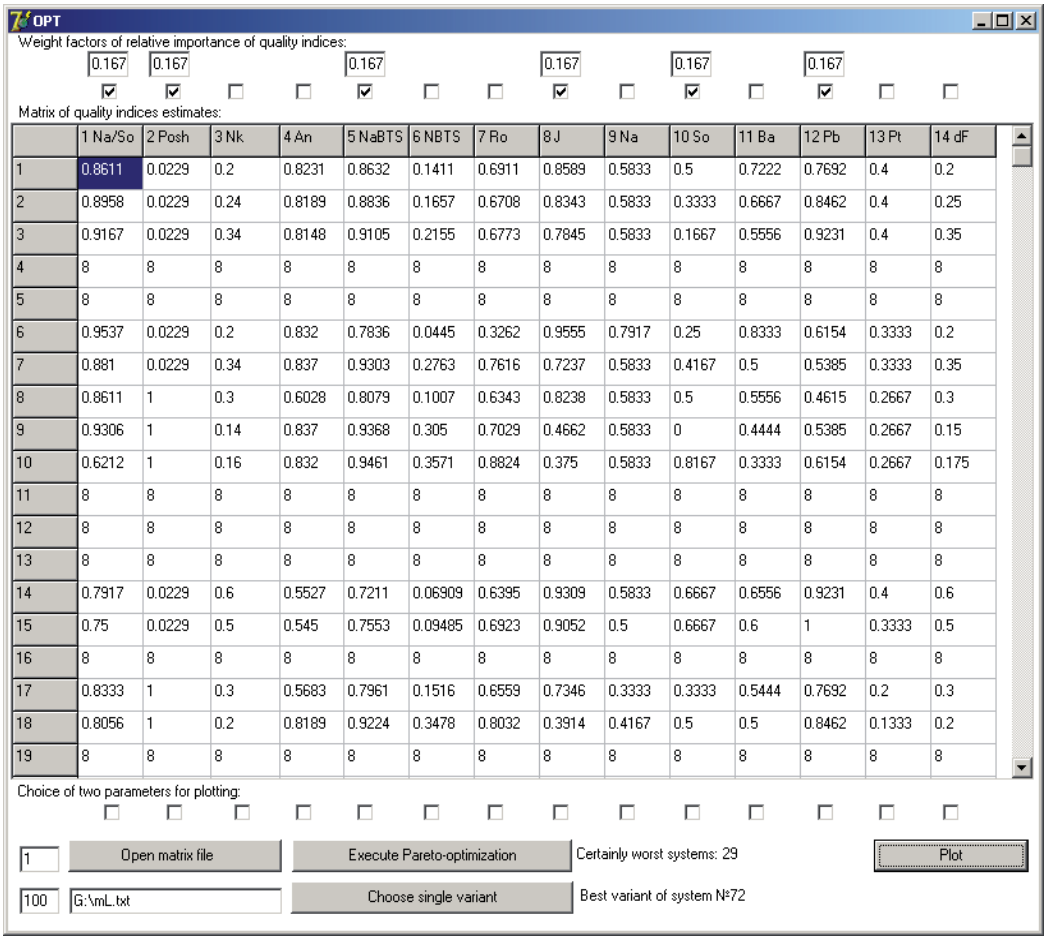
For a choice of optimal design solutions on the basis of multicriteria optimization methods, there was developed the program complex. It includes two parts solving the following issues.

1. Setting initial data and calculation of technical parameters for some permissible set of variants of CCN.
2. A choice of Pareto-optimal network variants and narrowing them to a single one.

Fig. 2 shows, as an example, the program complex interface. Here is shown part of table with values of 14 indicators for 19 CCN variants. There is the possibility to choose («tick off») concrete quality indicators to be taken into account at the multicriteria optimization. Besides, here are given values of coefficients of relative importance of chosen quality indicators.

There was selected a subset of Pareto-optimal variants including 71 network variants. Therewith 29 certainly worst variants are rejected. From the condition of minimum conditional criteria of preference as of the Pareto subset, a single variant is chosen (№72). It

is characterized by the following initial and calculated parameters: the number of subscribers is 30000; the area under service is 320 km²; activity of subscribers is 0.025 Erl; the frequency bandwidth is 4 MHz; the permissible probability of call blocking is 0.01; percentage of the connection quality deterioration time is 0.07; the density of service is 94 active subscribers per km²; the cluster size is 7; the number of base stations in the network is 133; the number of subscribers serviced by one BS is 226; the efficiency of radio frequency spectrum is $1.614 \cdot 10^{-4}$ active subscribers per Hz; the telephone load is 3.326 Erl; the probability of error is $5.277 \cdot 10^{-7}$; the angle of antenna radiation pattern is 120 degrees.



	1 Na/So	2 Posh	3 Nk	4 An	5 NaBTS	6 NBTS	7 Ro	8 J	9 Na	10 So	11 Ba	12 Pb	13 Pt	14 dF
1	0.8611	0.0229	0.2	0.8231	0.8632	0.1411	0.6911	0.8589	0.5833	0.5	0.7222	0.7692	0.4	0.2
2	0.8958	0.0229	0.24	0.8189	0.8836	0.1657	0.6708	0.8343	0.5833	0.3333	0.6667	0.8462	0.4	0.25
3	0.9167	0.0229	0.34	0.8148	0.9105	0.2155	0.6773	0.7845	0.5833	0.1667	0.5556	0.9231	0.4	0.35
4	8	8	8	8	8	8	8	8	8	8	8	8	8	8
5	8	8	8	8	8	8	8	8	8	8	8	8	8	8
6	0.9537	0.0229	0.2	0.832	0.7836	0.0445	0.3262	0.9555	0.7917	0.25	0.8333	0.6154	0.3333	0.2
7	0.881	0.0229	0.34	0.837	0.9303	0.2763	0.7616	0.7237	0.5833	0.4167	0.5	0.5385	0.3333	0.35
8	0.8611	1	0.3	0.6028	0.8079	0.1007	0.6343	0.8238	0.5833	0.5	0.5556	0.4615	0.2667	0.3
9	0.9306	1	0.14	0.837	0.9368	0.305	0.7029	0.4662	0.5833	0	0.4444	0.5385	0.2667	0.15
10	0.6212	1	0.16	0.832	0.9461	0.3571	0.8824	0.375	0.5833	0.8167	0.3333	0.6154	0.2667	0.175
11	8	8	8	8	8	8	8	8	8	8	8	8	8	8
12	8	8	8	8	8	8	8	8	8	8	8	8	8	8
13	8	8	8	8	8	8	8	8	8	8	8	8	8	8
14	0.7917	0.0229	0.6	0.5527	0.7211	0.06909	0.6395	0.9309	0.5833	0.6667	0.6556	0.9231	0.4	0.6
15	0.75	0.0229	0.5	0.545	0.7553	0.09485	0.6923	0.9052	0.5	0.6667	0.6	1	0.3333	0.5
16	8	8	8	8	8	8	8	8	8	8	8	8	8	8
17	0.8333	1	0.3	0.5683	0.7961	0.1516	0.6559	0.7346	0.3333	0.3333	0.5444	0.7692	0.2	0.3
18	0.8056	1	0.2	0.8189	0.9224	0.3478	0.8032	0.3914	0.4167	0.5	0.5	0.8462	0.1333	0.2
19	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Fig. 2. Interface of program complex.

As results of Pareto-optimization, there were obtained multivariate patterns of exchange (MPE) of the quality indicators, being of antagonistic character. For illustration, some MPE are presented at fig. 3. Each MPE point defines the potentially best values of each indicator which can be attained at fixed but arbitrary values of other quality indicators. MPE also show how the improvement of some quality indicators is achieved at the expense of other.

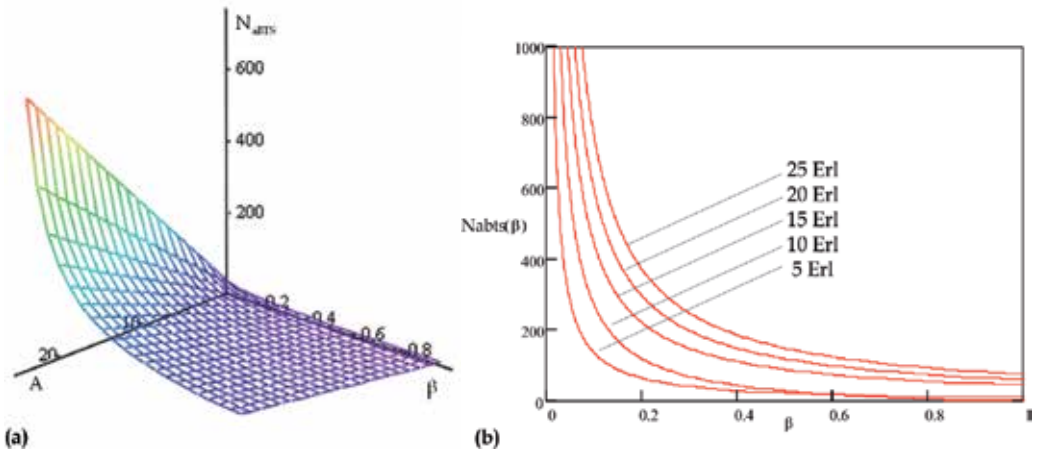


Fig. 3. MPE of the quality indicators (the number of subscribers serviced by one base station (a), the load, the activity of subscribers (b)) for CCN of GSM standard.

3.3 Features of a choice of Pareto-optimal routes

We have a set of permissible solutions (routes) on the finite network graph $G=(V,X)$, where $V=\{v\}$ – set of nodes, $E=\{e\}$ – set of network lines. Each route X is defined by a subset of the nodes and links. The goal task is presented by the model $\{X,F\}\to x^*$, where $X=\{x\}$ – set of permissible solutions (routes) on the network graph $G=(V,E)$; $F(x)$ – objective function of choice of the routes; x^* – optimal solution of the routing problem. The multicriteria approach of a choice of the best routes relies to perform decomposition of the function $F(x)$ to set (vector) partial choice functions. In this case on the set X it is given the vector of the objective function (Bezruk & Varich, 2011):

$$F(x)=(W_1(x),...,W_j(x),...,W_m(x)),$$

where components determine the values of quality routes indicators.

The route variant $x^*\in X$ is Pareto-optimal route if another route $x\in X$ doesn't exist, order to perform inequality $F_j(x^*)\le F_j(\tilde{x}), j=1,...,m,$ where at least one of the inequalities is strict. We propose to solve the problem of finding Pareto-optimal routes by using weight method. It is used for finding extreme values of the objective route function as a weighted sum of the partial choice functions for all possible values of the weighting coefficients λ_j:

$$\underset{\text{var}\,x\in X}{\text{extr}}=(F(x))=\sum_{j=1}^{m}\lambda_j W_j(x).$$

Pareto-optimal routes have some characteristic features. Particularly, Pareto-optimal alternative routes corresponds to the Pareto coordinated optimum partial objective functions $W_1(x),\ldots,W_j(x),\ldots,W_m(x)$. When selecting a subset of the Pareto-optimal routes there was dropped a certainly worst variant in terms of the absolute criteria of preference.

Pareto-optimal alternatives of the routes are equivalent to the Pareto criteria and could be used for organizing multipath routing in the multi-service telecommunication networks.

Network model consists of twelve nodes; they are linked by communication lines with losses (fig. 4).

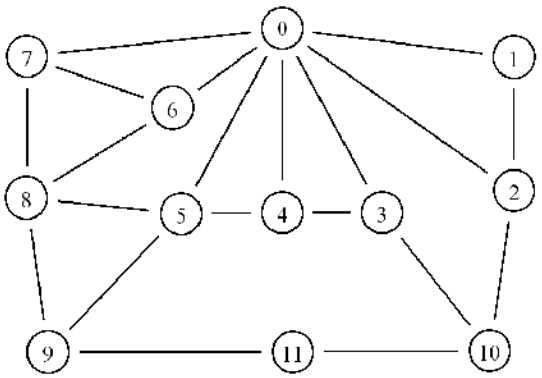


Fig. 4. The structure of the investigated network.

The quality indicators normalized to maximum values are presented in table 2.

The link	The delay time of packets transmission k_1	The level of packet loss k_2	The cost of using the line k_3
0-1	0.676	1	0.333
0-2	1	0.25	1
0-3	0.362	1	0.333
0-4	0.381	0.25	1
0-5	0.2	1	0.333
0-6	0.19	1	0.333
0-7	0.571	0.25	1
7-6	0.4	0.25	0.333
7-8	0.362	0.25	0.667
8-6	0.314	0.5	0.5
8-5	0.438	0.25	0.333
8-9	0.248	0.5	0.333
9-5	0.257	0.25	1
9-11	0.571	0.25	0.667
11-10	0.762	0.25	0.333
5-4	0.381	0.25	0.667
2-10	0.457	0.25	0.333
3-10	0.79	0.25	0.333
4-3	0.286	0.25	0.333
1-2	0.448	0.25	0.333

Table 2. Network quality indicators.

Network analysis shows that for each destination node there are many options to choose the route directly. For example, between node 0 and node 8 there are 22 routes.

Fig. 5 shows the set of the alternative routes between nodes 0 and 8 in the space of the quality indicators k_1 and k_2. Subset of the Pareto-optimal alternatives routes corresponds to the left lower border which includes three variants, they are marked (▲). This subset corresponds to be coordinated in Pareto optimum of the quality indicators.

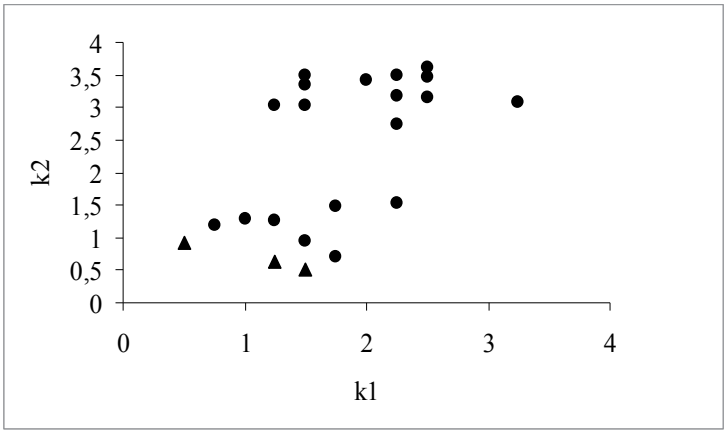


Fig. 5. Set of the routes between nodes 0 and 8.

The resulting subset of the Pareto-optimal alternative routes can be used for organizing multipath routing when using MPLS technology. It will allow to provide a load balancing and a traffic management and to provide given quality-of-service taking into account the set of the quality indicators.

3.4 Pareto-optimal choice of the speech codec

Proposed theoretical investigations can be used for Pareto-optimal choice of the speech codec used in IP-telephony systems (Bezruk & Skorik, 2010).

For carrying out the comparative analysis of basic speech codec and the optimal codec variant choice there have been used the data about 23 speech codecs described by the set of the technical and economic indicators: coding rate, quality of the speech coding, complexity of the realization, frame size, total time delay, etc. The initial values of the quality indicators are presented in table 3. It is easy to see that presented quality indicators are connected between each other with competing interconnections.

The time delay is increasing with frame size increasing as well as with complexity of the coding algorithm realization. Then, when transferring speech the permissible delay can not be bigger than 250 ms in one direction.

A frame size influences on the quality of a reproduced speech: the bigger is the frame, the more effective is the speech modeled. On other hand, the big frames increase an influence of the time delay on processing the information transferring. A frame size is defined by the compromise amongst these requirements.

№	Codec	Speech coding, kbps	Coding quality, MOS (1-5)	Complexity of the realization, MIPS	Frame size, ms	Total delay, ms
1	G 711	64	3,83	11,95	0,125	60
2	G 721	32	4,1	7,2	0,125	30
3	G 722	48	3,83	11,95	0,125	31,5
4	G 722(a)	56	4,5	11,95	0,125	31,5
5	G 722(b)	64	4,13	11,95	0,125	31,5
6	G 723.1(a)	5,3	3,6	16,5	30	37,5
7	G 723.1	6,4	3,9	16,9	30	37,5
8	G 726	24	3,7	9,6	0,125	30
9	G 726(a)	32	4,05	9,6	0,125	30
10	G 726(b)	40	3,9	9,6	0,125	30
11	G 727	24	3,7	9,9	0,125	30
12	G 727(a)	32	4,05	9,9	0,125	30
13	G 727(b)	40	3,9	9,9	0,125	30
14	G 728	16	4	25,5	0,625	30
15	G 729	8	4,05	22,5	10	35
16	G 729a	8	3,95	10,7	10	35
17	G 729b	8	4,05	23,2	10	35
18	G 729ab	8	3,95	11,5	10	35
19	G 729e	8	4,1	30	10	35
20	G 729e(a)	11,8	4,12	30	10	35
21	G 727(c)	16	4	9,9	0,125	30
22	G 728(a)	12,8	4,1	16	0,625	30
23	G 729d	6,4	4	20	10	35

Table 3. Codecs characteristics.

Complexity of the realization is connected with providing necessary calculations in real time. The coding algorithm complexity influences on the physical size of coding, decoding or combined devices, and also on its cost and power consumption.

In table 4 are presented some transformations results of the initial values of the quality indicators. In particular, there were performed the rationing operations of the indicators to their maximum values $k_{iн} = \frac{k_i}{k_{i\max}}$. These indicators were transformed to a comparable kind where all indicators had the same character depending on the technical codecs characteristics. In particular, for indicators k_{3n} and k_{5n} the transformations $k'_{3н} = \frac{1}{k_{3н}}$, $k'_{5н} = \frac{1}{k_{5н}}$ were done.

№	Codec	K_{1n}	K_{2n}	K'_{3n}	K_{4n}	K'_{5n}	Pareto-optimal choice
1	G 711	1	0,851	0,604	0,004	0,515	-
2	G 721	0,5	0,911	1	0,004	1	+
3	G 722	0,75	0,851	0,604	0,004	0,969	-
4	G 722(a)	0,875	1	0,604	0,004	0,969	+
5	G 722(b)	1	0,918	0,604	0,004	0,969	+
6	G 723.1(a)	0,083	0,8	0,439	1	0,818	+
7	G 723.1	0,1	0,867	0,424	1	0,818	+
8	G 726	0,375	0,822	0,748	0,004	1	-
9	G 726(a)	0,5	0,9	0,748	0,004	1	-
10	G 726(b)	0,625	0,866	0,748	0,004	1	+
11	G 727	0,375	0,822	0,727	0,004	1	-
12	G 727(a)	0,5	0,9	0,727	0,004	1	-
13	G 727(b)	0,625	0,866	0,727	0,004	1	-
14	G 728	0,25	0,889	0,281	0,021	1	+
15	G 729	0,125	0,9	0,317	0,333	0,879	+
16	G 729a	0,125	0,878	0,669	0,333	0,879	+
17	G 729b	0,125	0,9	0,309	0,333	0,879	-
18	G 729ab	0,125	0,878	0,626	0,333	0,879	-
19	G 729e	0,125	0,911	0,237	0,333	0,879	-
20	G 729e(a)	0,184	0,915	0,237	0,333	0,879	+
21	G 727(c)	0,25	0,889	0,727	0,004	1	-
22	G 728(a)	0,2	0,911	0,453	0,021	1	+
23	G 729d	0,1	0,889	0,359	0,333	0,879	+

Table 4. Transformed quality indicators.

On the base of received results there were considered the practical application features examined methods of the allocation of the Pareto-optimal speech codec variant set taking into account a set of the quality indicators as well as the unique design decision choice. From the initial set of the 23 speech codecs variants there was allocated the Pareto subset included 12 codecs variants (marked + in table 4).

The only one project decision was chosen from the condition of the scalar goal function extreme (9) with two different values of β defined characters of this function changing. In table 5 are presented the values of the given function for Pareto-optimal speech codecs variants at $\beta = 2$ and $\beta = 3$. It was obtained that an extreme goal function value, depending on β, is reached for the same speech codec G 722 (b).

Within statement of a problem we have chosen the codec of series G.722b which has following values of the quality indicators: speech coding – 64 kbps, coding quality – 4,13 MOS, complexity of the realization – 11,95 MIPS, the frame size – 0,125 ms, total delay – 31,5 ms.

№	Codec	Values $\xi_{\bar{k}}$ for diffrent β	
		$\beta = 2$	$\beta = 3$
2	G 721	0,35099	0,24688
4	G 722(a)	0,35039	0,28188
5	G 722(b)	0,35476	0,28532
6	G 723.1(a)	0,31677	0,25791
7	G 723.1	0,32312	0,26308
10	G 726(b)	0,32863	0,26445
14	G 728	0,27801	0,24056
15	G 729	0,26904	0,22785
16	G 729a	0,29103	0,23837
20	G 729e(a)	0,26912	0,22898
22	G 728(a)	0,28812	0,24582
23	G 729d	0,26927	0,22716

Table 5. Results of multicriteria optimization.

3.5 Network resources controlling

Let us consider some features of the short-term planning issues in the telecommunication system. There was shown the important place of multi-service network occupied with models, methods and facilities of network resources controlling in modern and perspective technologies. To the basic network resources facilities belong: channel resources control facilities (channels throughput, buffers size, etc), information resources control (user traffic).

Considered system was presented as the model of a distributed telecommunication system, consisting from a set of operating agents, for each autonomous system (fig 6).

In this model the process of network resources control was carried out by finding the distribution streams vector of the following type (Bezruk & Bukhanko, 2010):

$$\vec{K} = (k_1, k_2, ..., k_l), \ \sum_{i}^{l} k_i = 1,$$

with next limitation

$$0 \le k_i \le 1, \ i = \overline{1...l};$$

$$\lambda_i^{out} k_i \le c_i, \ i = \overline{1...l}.$$

Each element of this vector characterizes a part of outgoing user traffic from autonomous system operating agent transferred by using a corresponding channel. Within a given model, the task of network resources controlling comes to solving the optimization problem connected to function minimization.

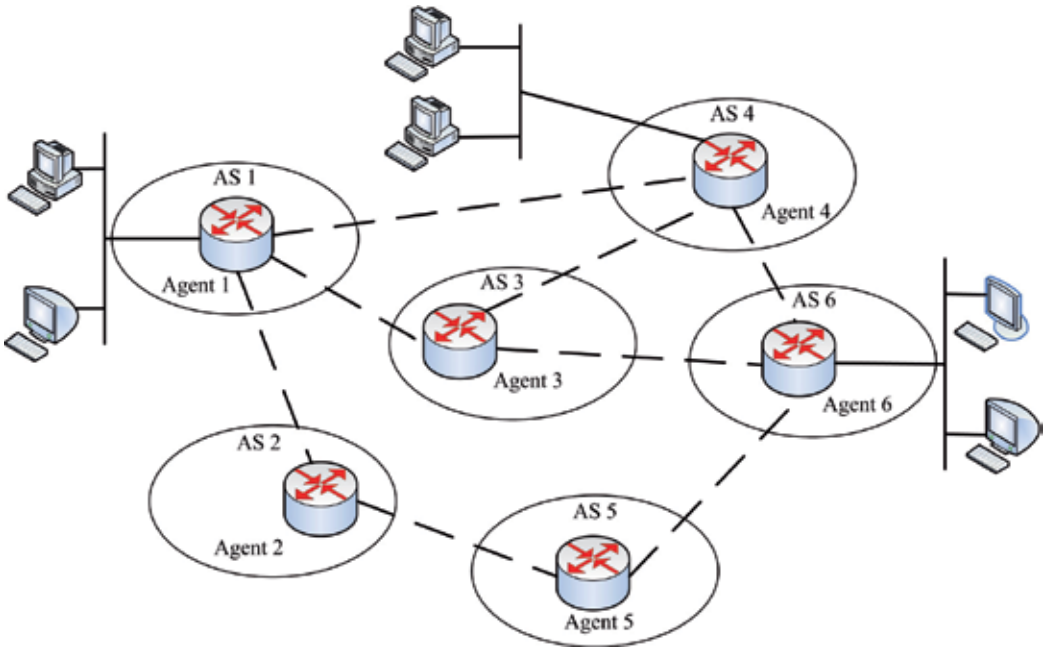


Fig. 6. Considered telecommunication system.

$$\varepsilon(\vec{K}) = \min(q_1\Phi + q_2\sigma_1(\vec{K}) + q_3\sigma_2(\vec{K})), \tag{10}$$

where $\sigma_1(\vec{K})$ – standard deviation of channels loading x_i, $i = \overline{1...l}$;

$$\sigma_1(\vec{K}) = \sqrt{\frac{1}{l-1}\sum_{i=1}^{l}\left(x_i - \bar{x}\right)^2};$$

$\sigma_2(\vec{K})$ – standard deviation of agents loading Z_i, $i = \overline{1...l}$;

$$\sigma_2(\vec{K}) = \sqrt{\frac{1}{l\text{-}1}\sum_{i=1}^{l}\left(Z_i - \bar{Z}\right)^2};$$

Φ – used routing protocol metric;

$$\Phi = \sum_{i=1}^{l}\varphi_i x_i;$$

φ_i – cost of full used channel ($\sum\lambda_i = c_i$);

q_1, q_2, q_3 – weight coefficients characterized the traffic balancing cost using standard metric, agents and channels loading.

The considered mathematical model of the distributed network resources controlling uses specific criteria of optimality included standard routing protocol metrics, a measure of channels and agents loading in given telecommunication network.

Obviously, under condition of $\sigma_1(\vec{K})$ and $\sigma_2(\overline{K})$ absence, function (10) becomes the model of the load balancing under the routes with equal or non-equal metric. However, absence of the decentralized control behind the autonomous system of telecommunication network can finally result in an uncontrollable overload. That fact is defined by the presence of additional minimized indicators leading to the practical value of the proposed model. Thus a choice of the relation of weight coefficients q_1, q_2 and q_3 is an independent problem demanding some future investigations and formalizations. In this model this task was dared with expert's estimations.

The proposed imitation model included up to 18 agents (fig. 7). Researches for different variants of connectivity between agents have been carried out.

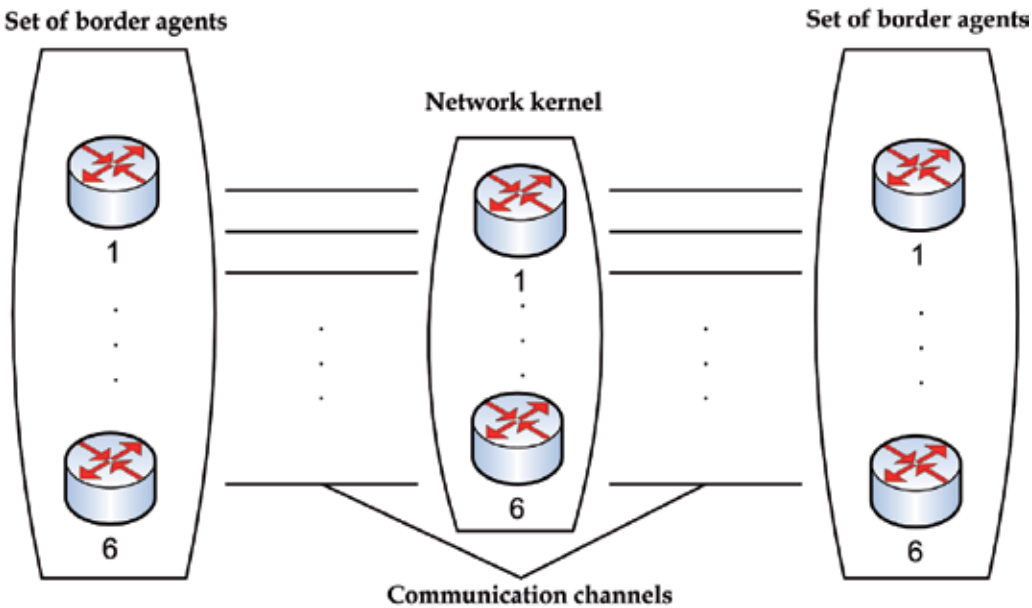


Fig. 7. Used imitation model.

During practical investigation there were analyzed several models of multipath routing and load balancing. These models are listed below:

M1 – model of routing by RIP;
M2 – model of multipath routing by an equal metric;
M3 – model of multipath routing by an non-equal metric (IGRP);
M4 – Gallagher stream model;
M5 – considered model with multicriteria account of two indicators (10);
M6 – considered model with multicriteria account of three indicators (10).

Below are presented some results of the analytic and imitation modeling within comparative analysis of considered existing and proposed models. These results are shown as dependences of the blocking probability and average delay time from the network loading (fig. 8).

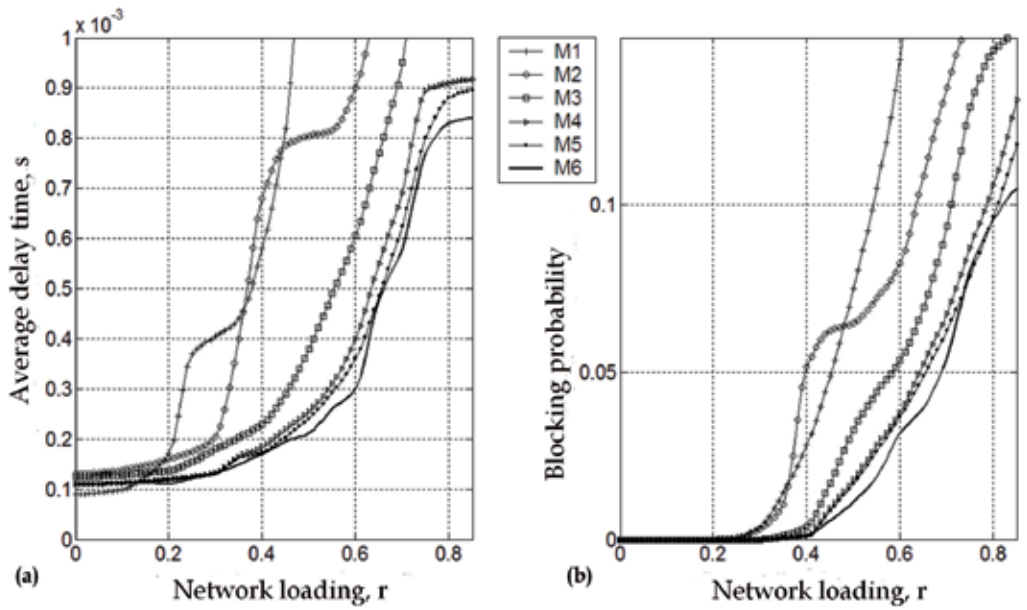


Fig. 8. Received dependences of average delay time (a) and blocking probability (b).

The use of the proposed models allows to:

- lower the average delay time (a) in comparison with the best known model (M4), for 3 – 12% (M5) and for 6 – 25% (M6);
- lower the general blocking probability (b) for 6 – 11% (M5) and 6 – 20% (M6).

4. Conclusion

The present work deals with the methodology of generating and selecting the variants of information systems when they are optimized in terms of the set of quality indicators. The multicriteria system-optimization problems are solved in three stages. By using the morphological approach a structural set of permissible variants of a system is initially generated. This set is mapped into the space of vector estimates. In this space a subset of Pareto-optimal estimates is selected, defining the potential characteristics of the system on the basis of the set of quality indicators. At the conclusive stage the only variant is selected amongst the Pareto-optimal variants of the system provided there exists an extreme of a certain scalar functional whose form is determined with the use of some additional information obtained from a customer.

Multicriteria optimization issues and methods based on Pareto conclusions are introduced for the long-term and short-term practical planning, designing and controlling within different types of telecommunication networks. In the process of solving the optimization problems we consider the set of network quality indicators as the different network topologies, transmission capacities of communication channels, various disciplines of service requests applied to different routing ways, etc.

Peculiarities of the long-term multicriteria optimization methods used for solving problems of the cellular networks planning are considered. As an example, the Pareto-optimization solution within planning of the cellular communication networks is also presented.

Practical features of the multicriteria approach in solving the optimal routing problem in the multi-service networks are considered within organizing multipath routing as well as speech codec choice based on a set of the quality indicators. The model of the information resources balancing on a basis of the decentralized operating agents system with a multicriteria account of chosen quality indicators is also offered. Considered adaptive balancing traffic algorithm improves the basic characteristics of the telecommunication network in a process of the short-term controlling for chosen cases of topologies.

5. Acknowledgment

The research described in this work was made possible in part by the scientific direction "Telecommunication and information networks optimization", headed by prof. Bezruk V., of the Communication Network Department within Kharkov National University of Radio Electronics, Ukraine.

6. References

Bezruk, V. & Skorik, Y. (2010). Optimization of speech codec on set of indicators of quality. *Proceedings of TCSET'2010 Modern problems of radio engineering, telecommunications and computer science,* p. 212, ISBN 978-966-553-875-2, Lviv – Slavske, Ukraine, February 23 – 27, 2010

Bezruk, V. & Bukhanko, O. (2010). Control mode of network resources in multiservice telecommunication systems on basis of distributed system of agents. *Proceedings of CriMiCo'2010 Microwave and Telecommunication Technology*, pp. 526-527, ISBN 978-966-335-329-6, Sevastopol, Crimea, Ukraine, September 13 – 17, 2010

Bezruk, V. & Varich, V. (2011). The multicriteria routing problem in multiservice networks with use composition quality indicators. *Proceedings of CriMiCo'2011 Microwave and Telecommunication Technology,* pp. 519 – 520, ISBN 978-966-335-254-8, Sevastopol, Crimea, Ukraine, September 12 – 16, 2011

Figueira, J. (Ed(s).). (2005). *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer Science + Business Media, Inc, ISBN 978-0-387-23081-8, Boston, USA

Saaty, P. (2005). *Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs and Risks,* RWS Publications, ISBN 1-888603-06-2, Pittsburgh, USA

Taha, H. (1997). *Operations Research: An Introduction*, Prentice Hall Inc., ISBN 0-13-272915-6, New Jersey, USA

Part 5

Traffic Engineering

12

Optical Burst-Switched Networks Exploiting Traffic Engineering in the Wavelength Domain

João Pedro[1,2] and João Pires[2]
[1]*Nokia Siemens Networks Portugal S.A.*
[2]*Instituto de Telecomunicações, Instituto Superior Técnico*
Portugal

1. Introduction

In order to simplify the design and operation of telecommunications networks, it is common to describe them in a layered structure constituted by a service network layer on top of a transport network layer. The service network layer provides services to its users, whereas the transport network layer comprises the infrastructure required to support the service networks. Hence, transport networks should be designed to be as independent as possible from the services supported, while providing functions such as transmission, multiplexing, routing, capacity provisioning, protection, and management. Typically, a transport network includes multiple network domains, such as access, aggregation, metropolitan and core, ordered by decreasing proximity to the end-users, increasing geographical coverage, and growing level of traffic aggregation.

Metropolitan and, particularly, core transport networks have to transfer large amounts of information over long distances, consequently demanding high capacity and reliable transport technologies. Multiplexing of lower data rate signals into higher data rate signals appropriate for transmission is one of the important tasks of transport networks. Time Division Multiplexing (TDM) is widely utilized in these networks and is the fundamental building block of the Synchronous Digital Hierarchy (SDH) / Synchronous Optical Network (SONET) technologies. The success of SDH/SONET is mostly due to the utilization of a common time reference, improving the cost-effectiveness of adding/extracting lower order signals from the multiplexed signal, the augmented reliability and interoperability, and the standardization of optical interfaces. SDH/SONET networks also generalized the use of optical fibre as the transmission medium of metropolitan and core networks. Essentially, when compared to twisted copper pair and coaxial cable, optical fibre benefits from a much larger bandwidth and lower attenuation, as well as being almost immune to electromagnetic interferences. These features are key to transmit information at larger bit rates over longer distances without signal regeneration.

Despite the proved merits of SDH/SONET systems, augmenting the capacity of transport networks via increasing their data rates is only cost-effective up to a certain extent, whereas

adding parallel systems by deploying additional fibres is very expensive. The prevailing solution to expand network capacity was to rely on Wavelength Division Multiplexing (WDM) to transmit parallel SDH/SONET signals in different wavelength channels of the same fibre. Nevertheless, since WDM was only used in point-to-point links, switching was performed in the electrical domain, demanding Optical-Electrical (OE) conversions at the input and Electrical-Optical (EO) conversions at the output of each intermediate node, as well as electrical switches. Both the OE and EO converters and the electrical switches are expensive and they represent a large share of the network cost.

Nowadays, transport networks already benefit from optical switching, thereby alleviating the use of expensive and power consuming OE and EO converters and electrical switching equipment operating at increasingly higher bit rates (Korotky, 2004). The main ingredients to support optical switching are the utilization of reconfigurable nodes, like Reconfigurable Optical Add/Drop Multiplexers (ROADMs) and Optical Cross-Connects (OXCs), along with a control plane, such as the Generalized Multi-Protocol Label Switching (GMPLS), (IETF, 2002), and the Automatically Switched Optical Network (ASON), (ITU-T, 2006). The control plane has the task of establishing/terminating optical paths (lightpaths) in response to connection requests from the service network. As a result, the current type of dynamic optical networks is designated as Optical Circuit Switching (OCS).

In an OCS network, bandwidth is allocated between two nodes by setting up one or more lightpaths (Zang et al., 2001). Consequently, the capacity made available for transmitting data from one node to the other can only be incremented or decremented in multiples of the wavelength capacity, which is typically large (e.g., 10 or 40 Gb/s). Moreover, the process of establishing a lightpath can be relatively slow, since it usually relies on two-way resource reservation mechanisms. Therefore, although the deployment of OCS networks only makes use of already mature optical technologies, these networks are inefficient in supporting bursty data traffic due to their coarse wavelength granularity and limited ability to adapt the allocated wavelength resources to the traffic demands in short time-scales, which can also increase the bandwidth waste due to capacity overprovisioning.

Diverse solutions have been proposed to overcome the limitations of OCS networks and improve the bandwidth utilization efficiency of future optical transport networks. The less disruptive approach consists of an optimized combination of optical and electrical switching at the network nodes. In this case, entire wavelength channels are switched optically at a node if the carried traffic flows, originated at upstream nodes, approximately occupy the entire wavelength capacity. Alternatively, traffic flows with small bandwidth requirements can be groomed (electrically) into one wavelength channel with enough spare capacity (Zhu et al., 2005). This hybrid switching solution demands costly OE/EO converters and electrical switches, albeit in/of smaller numbers/sizes than those needed in opaque implementations relying only on electrical switching. However, OCS networks with electrical grooming only become attractive when it is possible to estimate in advance the fractions of traffic to be groomed and switched transparently at each node, enabling to accurately dimension both the optical and electrical switches needed to accomplish an optimized trade-off between maximizing the bandwidth utilization and minimizing the electrical switching and OE/EO

conversion equipment. Otherwise, when the traffic pattern cannot be accurately predicted, this trade-off can become difficult to attain and both optical and electrical switches may have to be overdimensioned, hampering the cost-effectiveness of this hybrid approach.

The most advanced all-optical switching paradigm for supporting data traffic over optical transport networks is Optical Packet Switching (OPS). Ideally, OPS would replicate current store-and-forward packet-switched networks in the optical domain, thereby providing statistical multiplexing with packet granularity, rendering the highest bandwidth utilization when supporting bursty data traffic. In the full implementation of OPS, both data payload and their headers are processed and routed in the optical domain. However, the logical operations needed to perform address lookup are difficult to realize in the optical domain with state-of-the-art optics. Similarly to MPLS, Optical Label Switching (OLS) simplifies these logical operations through using label switching as the packet forwarding technique (Chang et al., 2006). In their simplest form, OPS networks can even rely on processing the header/label of each packet in the electrical domain, while the payload is kept in the optical domain. Nevertheless, despite the complexity differences of the implementations proposed in the literature, the deployment of any variant of OPS networks is always hampered by current limitations in optical processing technology, namely the absence of an optical equivalent of electronic Random-Access Memory (RAM), which is vital both for buffering packets while their header/label is being processed and for contention resolution (Tucker, 2006; Zhou & Yang, 2003), and the difficulty to fabricate large-sized fast optical switches, essential for per packet switching at high bit rates (Papadimitriou et al., 2003).

The above discussion highlighted that OCS networks are relatively simple to implement but inefficient for transporting bursty data traffic, whereas OPS networks are efficient for transporting this type of traffic but very difficult to implement with state-of-the-art optical technology. Next-generation optical networks would benefit from an optical switching approach whose bandwidth utilization and optical technology requirements lie between those of OCS and OPS. In order to address this challenge, an intermediate optical switching paradigm has been proposed and studied in the literature – Optical Burst Switching (OBS).

The basic premise of OBS is the development of a novel architecture for next-generation optical WDM networks characterized by enhanced flexibility to accommodate rapidly fluctuating traffic patterns without requiring major technological breakthroughs. A number of features have been identified as key to attain this objective (Chen et al., 2004). In order to overview some of them, consider an optical network comprising edge nodes, interfacing with the service network, and core nodes, as illustrated in Fig. 1. OBS networks grant intermediate switching granularity (between that of circuits and packets) via: assembling multiple packets into larger data containers, designated as data bursts, at the ingress edge nodes, enforcing per burst switching at the core nodes, and disassembling the packets at the egress edge nodes. Noteworthy, data bursts are only assembled and transmitted into the OBS network when data from the service network arrives at an edge node. This circumvents the stranded capacity problem of OCS networks, where the bandwidth requirements from the service network evolve throughout the lifetime of a lightpath and during periods of time can be considerably smaller than the provisioned capacity. Furthermore, the granularity at which the OBS network operates can be controlled through varying the number of packets contained in the data bursts, enabling to regulate the control and switching overhead.

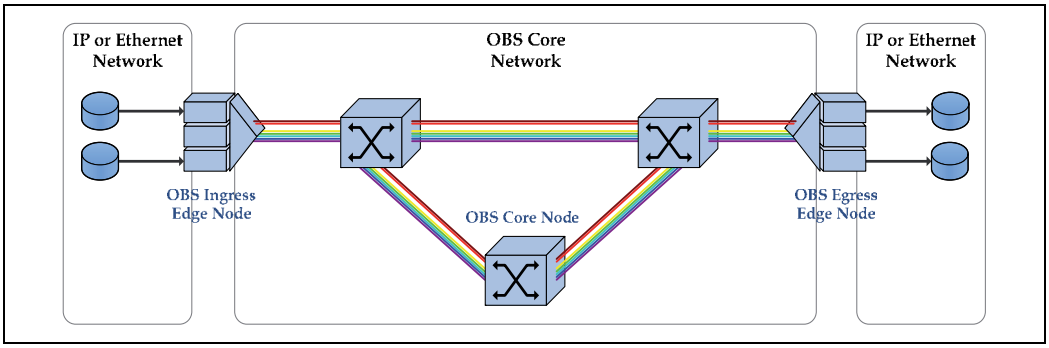


Fig. 1. Generic OBS network architecture.

In OBS networks, similarly to OCS networks, control information is transmitted in a separate wavelength channel and processed in the electronic domain at each node, avoiding complex optical processing functions inherent to OPS networks. More precisely, a data burst and its header packet are decoupled in both the wavelength and time domains, since they are transmitted in different wavelengths and the header precedes the data burst by an offset time. Channel separation of headers and data bursts, a distinctive feature of out-of-band signalling schemes, is suitable to efficiently support electronic processing of headers while preserving data in the optical domain, because OE/EO converters at the core nodes are only needed for the control channel. The offset time has a central role in OBS networks, since it is dimensioned to guarantee the burst header is processed and resources are reserved for the upcoming data burst before the latter arrives to the node. Accordingly, a data burst can cut through the core nodes all-optically, avoiding being buffered at their input during the time needed for header processing. Moreover, since the transmission of data bursts can be asynchronous, complex synchronization schemes are not mandatory. Combined, these features ensure OBS networks can be implemented without making use of optical buffering.

The prospects of deploying OBS in future transport networks can be improved provided that the bandwidth utilization achievable with OBS networks can be enhanced without significantly increasing their complexity or, alternatively, by easing their implementation without penalizing network performance. Noteworthy, OBS networks are technologically more demanding than OCS networks in several aspects. Firstly, although OBS protocols avoid optical buffering, OBS networks still demand some technology undergoing research, namely all-optical wavelength converters (Poustie, 2005) and fast optical switches scalable to large port counts (Papadimitriou et al., 2003). Secondly, the finer granularity of OBS is accomplished at the expense of a control plane more complex than the one needed for OCS networks (Barakat & Darcie, 2007). Nevertheless, the expected benefits of adopting a more bandwidth efficient optical switching paradigm fuelled significant research efforts in OBS, which even resulted in small network demonstrators (Sahara et al., 2003; Sun et al., 2005).

The performance of OBS networks is mainly limited by data loss due to contention for the same transmission resources between multiple data bursts (Chen et al., 2004). The lack of optical RAM limits the effectiveness of contention resolution in OBS networks. Wavelength conversion is usually assumed to be available to resolve contention for the same wavelength channel. In view of the complexity and immaturity of all-optical wavelength converters,

decreasing the number of converters utilized or using simpler ones without degrading performance would enhance the cost-effectiveness of OBS networks. Nevertheless, even if wavelength conversion is available, contention occurs when the number of bursts directed to the same link exceeds the number of wavelength channels. Moreover, the asynchronous transmission of data bursts creates voids between consecutive data bursts scheduled in the same wavelength channel, further contributing to contention. Consequently, minimizing these voids and smoothing burst traffic without resorting to complex contention resolution strategies would also improve the cost-effectiveness of OBS networks.

In alternative or as a complement to contention resolution strategies, such as wavelength conversion, the probability of resource contention in an OBS network can be proactively reduced using contention minimization strategies. Essentially, these strategies optimize the resources allocated for transmitting data bursts in such way that the probability of multiple data bursts contending for the same network resources is reduced. Contention minimization strategies for OBS networks mainly consist of optimizing the wavelength assignment at the ingress edge nodes to decrease contention for the same wavelength channel (Wang et al., 2003), mitigating the performance degradation from unused voids between consecutive data bursts scheduled in the same wavelength channel (Xiong et al., 2000), and selectively smoothing the burst traffic entering the network (Li & Qiao, 2004). Albeit the utilization of these strategies can entail additional network requirements, namely augmenting the (electronic) processing capacity in order to support more advanced algorithms, it is expected that the benefits in terms of performance or complexity reduction will justify their support.

This chapter details two contention minimization strategies, which when combined provide traffic engineering in the wavelength domain for OBS networks. The utilization of this approach is shown to significantly improve network performance and reduce the number of wavelength converters deployed at the network nodes, enhancing their cost-effectiveness.

The remaining of the chapter is organized as follows. The second section introduces the problem of wavelength assignment in OBS networks whose nodes have no wavelength converters or have a limited number of wavelength converters. A heuristic algorithm for optimizing the wavelength assignment in these networks is described and exemplified. The third section addresses the utilization of electronic buffering at the ingress edge nodes of OBS networks, highlighting its potential for smoothing the input burst traffic and describing how it can be combined with the heuristic algorithm detailed in the previous section to attain traffic engineering in the wavelength domain. The performance improvements and node complexity reduction made possible by employing these strategies in an OBS network are evaluated via network simulation in the fourth section. Finally, the fifth and last section presents the final remarks of the work presented in this chapter.

2. Priority-based wavelength assignment

OBS networks utilize one-way resource reservation, such as the Just Enough Time (JET) protocol (Qiao & Yoo, 1999). The principles of burst transmission are as follows. Upon assembling a data burst from multiple packets, the ingress node generates a Burst Header Packet (BHP) containing the offset time between itself and the data burst, as well as the length of the data burst. This node also sets a local timer to the value of the offset time.

The BHP is transmitted via a control wavelength channel and processed at the control unit of each node along the routing path of the burst. The control unit uses the information in the BHP to determine the resources (e.g., wavelength channel in the designated output fibre link) to be allocated to the data burst during the time interval it is expected to be traversing the core node. This corresponds to a delayed resource reservation, since the resources are not immediately set up, but instead are only set up just before the arrival time of the data burst. Furthermore, the resources are allocated to the burst during the time strictly necessary for it to successfully pass through the node. This minimizes the bandwidth waste because these resources can be allocated to other bursts in non-overlapping time intervals. Before forwarding the BHP to the next node, the control unit updates the offset time, reducing it by the amount of time spent by the BHP at the node. Meanwhile, the data burst buffered at the ingress node is transmitted after the timer set to the offset time expires. In case of successful resource reservation by its BHP at all the nodes of the routing path, the burst cuts through the core nodes in the optical domain until it arrives to the egress node. Otherwise, when resource reservation is unsuccessful at a node, both BHP and data burst are dropped at that node and the failed burst transmission is signalled to the ingress node.

As a result of using one-way resource reservation, there is a large probability that data bursts arrive at a core node on the same wavelength channel from different input fibre links and being directed to the same output fibre link of that node. This leads to contention for the same wavelength channel at the output fibre link. These contention events can be efficiently resolved using wavelength converters and/or minimized in advance through an optimized assignment of wavelengths at the ingress nodes. In view of the immaturity of all-optical wavelength converters, strategies for minimizing the probability of wavelength contention become of paramount importance in order to design cost-effective OBS core nodes.

2.1 Problem statement

Consider an OBS network modelled as a directed graph $G = (V, E)$, where $V = \{v_1, v_2, ..., v_N\}$ is the set of nodes, $E = \{e_1, e_2, ..., e_L\}$ is the set of unidirectional fibre links and the network has a total of N nodes and L fibre links. Each fibre link supports a set of W data wavelength channels, $\{\lambda_1, \lambda_2, \ldots, \lambda_{W-1}, \lambda_W\}$. Let $\Pi = \{\pi_1, \pi_2, \ldots, \pi_{|\Pi|-1}, \pi_{|\Pi|}\}$ denote the set of routing paths used to transmit data bursts in the network, E_i denote the set of fibre links traversed by path $\pi_i \in \Pi$, and γ_i denote the average traffic load offered to path π_i. It is assumed that the average offered traffic load values are obtained empirically or based on long-term predictions of the network load. Ideally, this input information would be used to formulate a combinatorial optimization problem for determining a wavelength search ordering, that is, an ordered list of all W wavelength channels, for each routing path such that a relevant performance metric, like the average burst blocking probability, is minimized. However, blocking probability performance metrics can only be computed via network simulation or, in particular cases, estimated by solving a set of non-linear equations (Pedro et al., 2006a). As a result, the objective function cannot be expressed in terms of the problem variables in an analytical closed-form manner (Teng & Rouskas, 2005). Moreover, even if this was possible, the size of the solutions search space would grow steeply with the number of wavelength channels W and the number of routing paths

$|\Pi|$, since there are $(W!)^{|\Pi|}$ combinations of wavelength channel orderings. Consequently, for OBS networks of realistic size, this would prevent computing the optimum wavelength search orderings in a reasonable amount of time.

In view of the aforementioned limitations in both problem formulation and resolution, the wavelength search orderings must be computed without knowing the resulting average burst blocking probability and by relying on heuristic algorithms. Notably, when the core nodes have limited or no wavelength conversion capabilities, burst blocking probability is closely related with the expected amount of unresolved wavelength contention. Consider two routing paths, π_1 and π_2, that traverse a common fibre link. Clearly, the chances of data bursts going through these paths and contending for the same wavelength channel at the common fibre link are minimized if their ingress nodes search for an available wavelength using opposite orderings of the wavelengths, that is, the ingress node of π_1 uses, for instance, λ_1, λ_2, …, λ_{W-1}, λ_W, whereas the ingress node of π_2 uses λ_W, λ_{W-1}, …, λ_2, λ_1. This simple scenario is illustrated is Fig. 2 for $W = 4$, where most of the burst traffic on π_1 (π_2) will go through λ_1, λ_2 (λ_4, λ_3). However, in realistic network scenarios, each routing path shares fibre links with several other paths and, consequently, it is not feasible to have opposite wavelength search orderings for each pair of overlapping paths. Still, as long as it is possible for two overlapping paths to have two different wavelength channels ranked as the highest priority wavelengths, the probability of wavelength contention among data bursts going through these paths is expected to be reduced. This observation constitutes the foundation of the heuristic traffic engineering approaches described in the following.

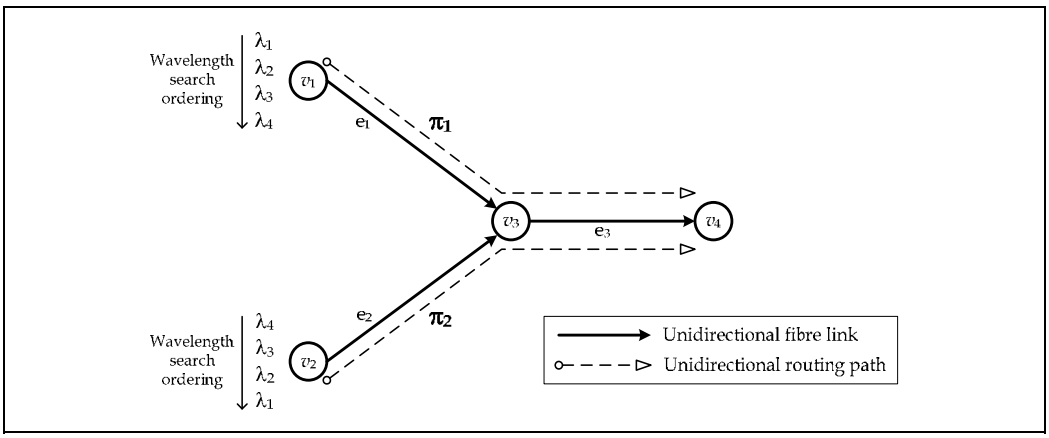


Fig. 2. Example OBS network with opposite wavelength search orderings.

2.2 Heuristic minimum priority interference

Intuitively, the chances of wavelength contention between data bursts going through different routing paths are expected to increase with both the average traffic load offered to the paths and with the number of common fibre links. Bearing this in mind, it is useful to define the concept of interference level of routing path π_i on routing path π_j with $i \neq j$ as,

$$I(\pi_i, \pi_j) = \gamma_i |E_i \cap E_j|, \quad (1)$$

where $|E_i \cap E_j|$ denotes the number of fibre links shared by both paths, and to define the combined interference level between routing paths π_i and π_j with $i \neq j$ as,

$$I^c(\pi_i, \pi_j) = I(\pi_i, \pi_j) + I(\pi_j, \pi_i) = (\gamma_i + \gamma_j)\,|E_i \cap E_j|\,. \tag{2}$$

The higher the combined interference level between two routing paths, the higher the likelihood that data bursts going through those paths will contend for the same fibre link resources. Consequently, routing paths with higher combined interference level should use wavelength search orderings as opposed as possible. This constitutes the basic principle exploited by First Fit-Traffic Engineering (FF-TE) (Teng & Rouskas, 2005), which was the first offline algorithm proposed to determine wavelength search orderings that are expected to reduce the probability of wavelength contention. However, this algorithm oversimplifies the problem resolution by computing a single wavelength search ordering for all the routing paths with the same ingress node. A detailed discussion of the limitations of the FF-TE algorithm is presented in (Pedro et al., 2006b). To overcome these shortcomings, the more advanced Heuristic Minimum Priority Interference (HMPI) algorithm, which computes an individual wavelength search ordering per routing path, is described below.

2.2.1 Algorithm description

The algorithm proposed in (Pedro et al., 2006b) for minimizing wavelength contention aims to determine an individual wavelength search ordering for each routing path with a reduced computational effort. The HMPI algorithm uses as input information the network topology, the routing paths and the average traffic load offered to the routing paths.

In order to determine the wavelength search ordering of a routing path, a unique priority must be assigned to each of the wavelengths. The wavelength ranked with the highest priority, called the primary wavelength, is expected to carry the largest amount of burst traffic going through the routing path. The other wavelengths, ordered by decreasing priority, expectedly carry diminishing amounts of burst traffic. In view of the importance of the primary wavelengths, the HMPI algorithm comprises a first stage dedicated to optimize them, consisting of the following three steps.

(**S1**) Reorder the routing paths of Π such that if $i < j$ one of the following conditions holds,

$$\sum_{\substack{\pi_k \in \Pi, \\ k \neq i}} I(\pi_i, \pi_k) > \sum_{\substack{\pi_k \in \Pi, \\ k \neq j}} I(\pi_j, \pi_k)\,; \tag{3}$$

$$\sum_{\substack{\pi_k \in \Pi, \\ k \neq i}} I(\pi_i, \pi_k) = \sum_{\substack{\pi_k \in \Pi, \\ k \neq j}} I(\pi_j, \pi_k) \text{ and } |E_i| > |E_j|\,. \tag{4}$$

(**S2**) Consider W sub-sets of the routing paths, one per wavelength, initially empty, that is, $|\Pi_j| = 0$ for $j = 1, \ldots, W$. Following the routing path ordering defined for Π, include path π_i in the sub-set Π_j such that for any $k \neq j$ one of the subsequent conditions holds,

$$\sum_{\substack{\pi_l \in \Pi_j, \\ l \neq i}} I^c(\pi_i, \pi_l) < \sum_{\substack{\pi_l \in \Pi_k, \\ l \neq i}} I^c(\pi_i, \pi_l)\,; \tag{5}$$

$$\sum_{\substack{\pi_l \in \Pi_j, \\ l \neq i}} I^c(\pi_i, \pi_l) = \sum_{\substack{\pi_l \in \Pi_k, \\ l \neq i}} I^c(\pi_i, \pi_l) \text{ and } |\Pi_j| > |\Pi_k|. \quad (6)$$

(**S3**) Select wavelength channel λ_j as the primary wavelength of all the paths in sub-set Π_j, that is,

$$P(\lambda_j, \pi_i) = \begin{cases} W, \text{ if } \pi_i \in \Pi_j \\ 0, \text{ otherwise} \end{cases}. \quad (7)$$

The first step of this stage of the HMPI algorithm is used to order the routing paths by decreasing interference level on the remaining paths. Ties are broken by giving preference to the longer routing paths. Considering W sub-sets of routing paths, the second step sequentially includes each routing path on the sub-set with minimum combined interference level between the routing path and the paths already included in the sub-set. Ties are broken by preferring the sub-set with larger number of paths. Finally, the third step assigns to all routing paths of a sub-set the primary wavelength associated with that sub-set. As a result of this stage, the routing paths with minimum combined interference level, carrying data bursts that are less prone to contend with each other for the same wavelength channel, will share the same primary wavelength.

In the second stage of the algorithm, the non-primary wavelengths for all routing paths are determined sequentially, starting with the second preferred wavelength channel and ending with the least preferred wavelength. When determining for each routing path the wavelength with priority $p < W$, it is intuitive to select one to which has been assigned, so far in the algorithm execution, the lowest priorities on routing paths that share fibre links with the routing path being considered. This constitutes the basic rule used in the second stage of the HMPI algorithm.

The following steps are executed for priorities $1 \leq p \leq W - 1$ in decreasing order and considering, for each priority p, all the routing paths according to the path ordering defined in the first stage of the algorithm.

(**S1**) Let $\Lambda = \{\lambda_j : P(\lambda_j, \pi_i) = 0, 1 \leq j \leq W\}$ denote the initial set of candidate wavelengths, containing all wavelengths that have been assigned a priority of zero on routing path π_i. If $|\Lambda| = 1$, go to (**S7**).

(**S2**) Let $\mathrm{P} = \{k : \exists \pi_l, l \neq i, P(\lambda_j, \pi_l) = k, |E_l \cap E_i| > 0, \lambda_j \in \Lambda\}$ be the set of priorities that have already been assigned to candidate wavelengths on paths that overlap with π_i.

(**S3**) Let $\psi = \min_{\lambda_j \in \Lambda} \max_{\pi_l \in \Pi} \{P(\lambda_j, \pi_l) : l \neq i, |E_l \cap E_i| > 0, P(\lambda_j, \pi_l) \in \mathrm{P}\}$ be the lowest priority among the set containing the highest priority assigned to each candidate wavelength on paths that share links with π_i. Update the set of candidate wavelengths as follows,

$$\Lambda \leftarrow \Lambda \setminus \{\lambda_j : \exists \pi_l, l \neq i, P(\lambda_j, \pi_l) > \psi, |E_l \cap E_i| > 0\}; \quad (8)$$

If $|\Lambda| = 1$, go to (**S7**).

(**S4**) Define $C(\lambda_j, e_m) = \sum\{\gamma_l : E_l \supset e_m, |E_l \cap E_i| > 0, P(\lambda_j, \pi_l) = \psi\}$ as the cost associated with wavelength channel $\lambda_j \in \Lambda$ on link $e_m \in E_i$ and $\alpha_e = \min_{\lambda_j \in \Lambda} \max_{e_m \in E_i} C(\lambda_j, e_m)$ as the minimum cost among the set containing the highest cost associated with each candidate wavelength on the fibre links of π_i. Update the set of candidate wavelengths as follows,

$$\Lambda \leftarrow \Lambda \setminus \{\lambda_j : \exists e_m, C(\lambda_j, e_m) > \alpha_e, e_m \in E_i\}; \quad (9)$$

If $|\Lambda| = 1$, go to (**S7**).

(**S5**) Define $C(\lambda_j, \pi_i) = \sum_{e_m \in E_i} C(\lambda_j, e_m)$ as the cost associated with wavelength λ_j on path π_i and $\alpha_\pi = \min_{\lambda_j \in \Lambda} C(\lambda_j, \pi_i)$ as the minimum cost among the costs associated with the candidate wavelengths on π_i. Update the set of candidate wavelengths as follows,

$$\Lambda \leftarrow \Lambda \setminus \{\lambda_j : C(\lambda_j, \pi_i) > \alpha_\pi\}; \quad (10)$$

If $|\Lambda| = 1$, go to (**S7**).

(**S6**) Update the set of priorities assigned to the candidate wavelengths as follows,

$$P \leftarrow P \setminus \{k : k \geq \psi\}; \quad (11)$$

If $|P| > 0$, go to (**S3**). Else, randomly select a candidate wavelength $\lambda \in \Lambda$.

(**S7**) Assign priority p to the candidate wavelength $\lambda \in \Lambda$ on path π_i, that is, $P(\lambda, \pi_i) = p$.

The first step of the second stage of the HMPI algorithm is used to define the candidate wavelength channels by excluding the ones that have already been assigned a priority larger than zero on the routing path, whereas the second step determines the priorities assigned to these wavelengths on paths that overlap with the routing path under consideration. The third, fourth and fifth step are used to reduce the number of candidate wavelengths. As soon as there is only one candidate wavelength, it is assigned to it the priority p on path π_i, concluding the iteration. In the third step, the highest priority already assigned to each of the candidate wavelength channels on paths that overlap with π_i is determined. Only the wavelengths with the lowest of these priorities are kept in the set of candidates. If needed, the fourth step tries to break ties by associating a cost with each candidate wavelength on each fibre link of π_i. This cost is given by the sum of the average traffic load offered to paths that traverse the fibre link and use the wavelength with priority ψ. The wavelengths whose largest link cost, among all links of π_i, is the smallest one (α_e) are kept as candidates. When there are still multiple candidate wavelengths, the fifth step associates a cost with each wavelength on path π_i, which is simply given by the sum of the cost associated to the wavelength on all links of the routing path. The candidate wavelengths with smallest path cost (α_π) are kept. If necessary, the sixth step removes the priorities equal or larger than ψ from the set of priorities assigned to candidate wavelengths on paths that overlap with the path being considered and repeats the iteration. Finally, if all priorities have been removed and there are still multiple candidate wavelengths, one of them is randomly selected.

As the outcome of executing the HMPI algorithm, each wavelength channel λ_j is assigned a unique priority on routing path π_i, $1 \leq P(\lambda_j, \pi_i) \leq W$. Equivalently, this solution for the priority assignment problem can be represented as an ordering of the W wavelengths, $\{\lambda^1(\pi_i), \lambda^2(\pi_i), \ldots, \lambda^j(\pi_i), \ldots, \lambda^W(\pi_i)\}$, where $\lambda^j(\pi_i)$ denotes the j^{th} wavelength channel to be searched when assigning a wavelength to data bursts directed to routing path π_i. In order to enforce these search orderings, each of these lists must be uploaded from the point where they are computed to the ingress nodes of the routing paths. Hence, assuming single-path routing, each ingress node will have to maintain at most $N - 1$ lists of ordered wavelengths.

The computational complexity of the HMPI algorithm, as derived in (Pedro et al., 2009c), is given by $O(W^2 \cdot |\Pi|^2)$, that is, in the worst case it scales with the square of the number of wavelength channels times the square of the number of routing paths.

2.2.2 Illustrative example

In order to give a better insight into the HMPI algorithm, consider the example OBS network of Fig. 3, which has 6 nodes and 8 fibre links (Pedro et al., 2009c). The number of routing paths used to transmit bursts in the network is $|\Pi| = 6$ and each fibre link supports a number of wavelength channels $W = 4$. Moreover, the average traffic load offered to each routing path is 1, except for routing path π_4, which has an average offered traffic load of 1.2, that is, $\gamma_i = 1$ for $i = 1, 2, 3, 5, 6$ and $\gamma_4 = 1.2$.

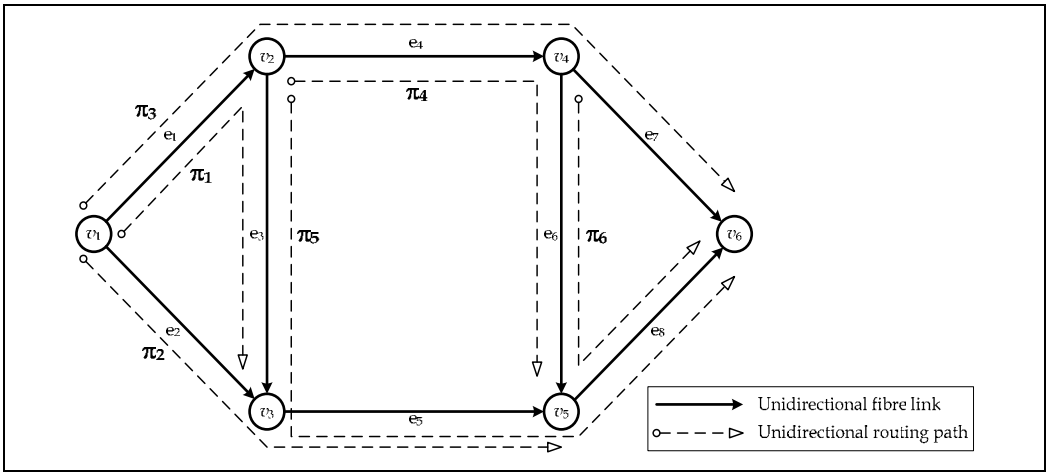


Fig. 3. OBS network used to exemplify the HMPI algorithm (Pedro et al., 2009c).

The HMPI algorithm starts by computing the interference level of all pairs of routing paths, as shown in Table 1. Step (**S1**) of the first stage of the algorithm orders the routing paths by decreasing order of their interference level over other paths, which results in the path order $\{\pi_5, \pi_4, \pi_3, \pi_1, \pi_6, \pi_2\}$. The path with the highest interference level over other paths is π_5, which overlaps with three paths, and the path with the second highest interference level over other paths is π_4, which overlaps with two paths. Although π_3, π_1 and π_6 also overlap with two paths, π_4 is offered more traffic load and consequently can cause more contention. In addition, π_3 precedes π_1 and π_6 because it is longer than the later paths. Since paths π_1 and

π_6 are tied, the path with the smallest index was given preference. Finally, the path with the lowest interference level over other paths is π_2.

$I(\pi_i, \pi_j)$	π_1	π_2	π_3	π_4	π_5	π_6
π_1	--	0	1	0	1	0
π_2	0	--	0	0	1	0
π_3	1	0	--	1	0	0
π_4	0	0	1.2	--	0	1.2
π_5	1	1	0	0	--	1
π_6	0	0	0	1	1	--

Table 1. Interference level of the routing paths.

Step (**S2**) starts by creating one sub-set of routing paths per wavelength, that is, Π_1, Π_2, Π_3, Π_4. Following the determined path order, π_5 is included in the first empty sub-set, Π_1. Path π_4 is also included in Π_1, because $I^C(\pi_4, \pi_5) = 0$ and Π_1 has more paths than the remaining sub-sets. Since path π_3 overlaps with π_4, $I^C(\pi_3, \pi_4) = 2.2$, and π_4 is already included in Π_1, π_3 is included in the empty sub-set Π_2. Moreover, path π_1 overlaps with both π_5 and π_3 and thus it is included in empty sub-set Π_3. Path π_6 can be included in sub-sets Π_2 and Π_3 because it only overlaps with the paths of Π_1. The tie is broken by selecting the sub-set with smallest index, that is, Π_2. Similarly, path π_2 is also included in this sub-set as it does not overlap with the paths in Π_2 and Π_3 and $|\Pi_2| > |\Pi_3|$. Since every path has been included in one sub-set, $\Pi_1 = \{\pi_4, \pi_5\}$, $\Pi_2 = \{\pi_2, \pi_3, \pi_6\}$ and $\Pi_3 = \{\pi_1\}$, step (**S3**) concludes the first stage of the algorithm by making λ_1 the primary wavelength of paths π_4 and π_5, λ_2 the primary wavelength of paths π_2, π_3 and π_6, and λ_3 the primary wavelength of path π_1. The other wavelengths are temporarily assigned priority 0 on the routing paths. Table 2 shows the priorities assigned to the wavelengths on the routing paths after the entire HMPI algorithm has been executed.

$P(\lambda_j, \pi_i)$	π_1	π_2	π_3	π_4	π_5	π_6
λ_1	1	1	1	4	4	1
λ_2	2	4	4	1	1	4
λ_3	4	3	2	2	2	3
λ_4	3	2	3	3	3	2

Table 2. Wavelengths priority on the routing paths.

The second stage of the algorithm is initiated with $p = 3$ and proceeds path by path according to the order already defined. For path π_5, the algorithm starts by creating the initial set of candidate wavelengths, $\Lambda = \{\lambda_2, \lambda_3, \lambda_4\}$, in (**S1**). Since this path overlaps with π_1, π_2 and π_6, the set of priorities assigned to wavelengths of Λ on these paths, determined in (**S2**), is $P = \{0, 4\}$. Wavelength λ_4 is assigned priority 0 on all paths that overlap with π_5 and thus $\rho = 0$. Accordingly, in (**S3**) the set of candidate wavelengths is updated, $\Lambda = \{\lambda_4\}$, and λ_4 is assigned priority 3 on path π_5. For path π_4, $\Lambda = \{\lambda_2, \lambda_3, \lambda_4\}$, $P = \{0, 4\}$, and $\rho = 0$. The set of

candidate wavelengths is updated to $\Lambda = \{\lambda_3, \lambda_4\}$, because both λ_3 and λ_4 are assigned priority 0 on paths that overlap with π_4. In this particular case, the algorithm cannot break the tie and in (**S7**) randomly selects wavelength λ_4 to be assigned priority 3 on path π_4. For the remaining paths, there is only one candidate wavelength whose priority on other paths equals ρ. Wavelength λ_4 is assigned priority 3 on paths π_3 and π_1 and wavelength λ_3 is assigned this priority on paths π_6 and π_2.

The second stage of the algorithm is executed again, but with $p = 2$. For path π_5, the initial set of candidate wavelengths is $\Lambda = \{\lambda_2, \lambda_3\}$. Both wavelengths are assigned priority 4 on at least one of the paths that overlaps with π_5 ($\rho = 4$), λ_2 on π_2 and π_6 and λ_3 on π_1. Paths π_1, π_2, and π_6 share with π_5 links e_3, e_5 and e_8, respectively, and the average traffic load offered to these paths is 1. Thus, according to (**S4**), the cost associated with λ_2 and λ_3 on each link is at most 1 ($\alpha_e = 1$). However, λ_2 has this link cost on two links, which in (**S5**) results in a cost $C(\lambda_2, \pi_5) = 2$, whereas λ_3 has this link cost on a single link, $C(\lambda_3, \pi_5) = 1$. Consequently, $\alpha_\pi = 1$ and the set of candidate wavelengths is updated to $\Lambda = \{\lambda_3\}$. For path π_4, $\Lambda = \{\lambda_2, \lambda_3\}$, $P = \{0, 3, 4\}$, and $\rho = 3$. Only wavelength λ_3 is used with a priority smaller or equal than 3 in all links, which reduces the set of candidates to λ_3. In the case of path π_3, $\Lambda = \{\lambda_1, \lambda_3\}$ and λ_1 is assigned priority 4 on π_4, whereas λ_3 is assigned this priority on π_1. Since $\gamma_4 > \gamma_1$, the highest link cost associated to λ_1 is larger than that for λ_3, and the candidate wavelengths are reduced to λ_3. For path π_1, $\Lambda = \{\lambda_1, \lambda_2\}$ and both these wavelengths observe $\rho = 4$, $\alpha_e = 1$ and $\alpha_\pi = 1$. The algorithm has to randomly select one of the wavelengths (λ_2). For both π_6 and π_2, $\Lambda = \{\lambda_1, \lambda_4\}$, $\rho = 3$, but only λ_4 is assigned a priority smaller or equal to 3 in all of the links. The set of candidate wavelengths is reduced to $\Lambda = \{\lambda_4\}$.

Finally, for $p = 1$ the wavelength assignment is trivial, because there is only one wavelength still assigned priority 0 on each path. The complete wavelength search ordering of each path can be obtained from Table 2. The following observations show that these orderings should effectively reduce contention. Firstly, overlapping paths do not share the same primary wavelength. Instead, primary wavelengths are reused by link-disjoint routing paths (e.g., λ_2 is the primary wavelength of π_2, π_3 and π_6). Secondly, paths use with smallest possible priority the primary wavelengths of overlapping paths (e.g., π_1, π_2 and π_6 overlap with π_5 and use the primary wavelength of this path with priority 1).

3. Traffic engineering in the wavelength domain

Noteworthy, at the ingress edge nodes of an OBS network, data bursts are kept in electronic buffers before a wavelength channel is assigned to them and they are transmitted optically towards the egress edge nodes. Clearly, the flexibility of scheduling data bursts in the wavelength channels is considerably higher when the bursts are still buffered at the ingress nodes than when they have already been converted to the optical domain. For instance, a data burst can be delayed at one of the ingress buffers by the exact amount of time required for a wavelength channel to become available in the designated output fibre link. This procedure is not possible at the core nodes due to the lack of optical RAM. The capability of delaying data bursts at an ingress node by a random amount of time, not only increases the chances of successfully scheduling bursts at the output fibre link of their ingress nodes, but also enables implementing strategies that reduce in advance the probability of contention at the core nodes.

The Burst Overlap Reduction Algorithm proposed in (Li & Qiao, 2004) exploits the additional degree of freedom provided by delaying data bursts at the electronic buffers of the ingress nodes to shape the burst traffic departing from these nodes in such way that the probability of contention at the core nodes can be reduced. The principle underlying BORA is that a decrease on the number of different wavelength channels allocated to the data bursts assembled at an ingress node can smooth the burst traffic at the input fibre links of the core nodes and, as a result, reduce the probability that the number of overlapping data bursts directed to the same output fibre link exceeds the number of wavelength channels. In its simpler implementation, BORA relies on using the same wavelength search ordering at all the ingress nodes of the network and utilizing the buffers in these nodes to transmit the maximum number of bursts in the first wavelength channels according to such ordering. In order to limit the extra transfer delay incurred by data bursts, as well as the added buffering and processing requirements, the ingress node can impose a maximum ingress burst delay, $\Delta t_{\max}^{\mathrm{RAM}}$, defined as the maximum amount of time a data burst can be kept at an electronic buffer of its ingress node excluding the time required to assemble the burst and the offset time between the data burst and its correspondent BHP.

The concept of BORA is appealing in OBS networks with wavelength conversion, since these algorithms have not been designed to mitigate wavelength contention. Moreover, BORA algorithms do not account for the capacity fragmentation of the wavelength channels, which is also a performance limiting factor in OBS networks. These limitations have motivated the development of a novel strategy in (Pedro et al., 2009b) that also exploits the electronic buffers of the ingress edge nodes to selectively delay data bursts, while providing a twofold advantage over BORA: enhanced contention minimization at the core nodes and support of core node architectures with relaxed wavelength conversion capabilities.

The first principle of the proposed strategy is related with the availability of RAM at the ingress nodes. In the process of judiciously delaying bursts to schedule them using the smallest number of different wavelength channels, the delayed bursts can be scheduled with minimum voids between them and the preceding bursts already scheduled on the same wavelength channel. This is only possible because the bursts assembled at the node can be delayed by a random amount of time. The serialization of data bursts not only smoothes the burst traffic, with the consequent decrease of the chances of contention at the core nodes, but also reduces the fragmentation of the wavelengths capacity at the output fibre links of the ingress nodes. These serialized data bursts traverse the core nodes, where some of them must be converted to other wavelength channels to resolve contention. The wavelength conversions break the series of data bursts and, as a result, create voids between a burst converted to another wavelength channel and the bursts already scheduled on this wavelength. A large number of these voids lead to wasting bandwidth, as the core nodes will not be able to use them to carry data.

In essence, the first key principle consists of serializing data bursts at the ingress nodes to mitigate the voids between them. Noticeably, if these bursts traverse a set of common fibre links without experiencing wavelength conversion, the formation of unusable voids is reduced at those links. Hence, the second key principle of the proposed strategy consists of improving the probability that serialized bursts routed via the same path are kept in the same wavelength channel for as long as possible. This can reduce the number of unusable

voids created in the fibre links traversed before wavelength conversion is used, improving network performance.

The task of keeping the data bursts, which are directed to the same routing path and have been serialized at the ingress node, in the same wavelength channel requires minimizing the chances that bursts on overlapping routing paths contend for the same wavelength channel and, as a result, demand wavelength conversion. This objective is the same as that of the HMPI algorithm presented in Section 2. For that reason, the strategy proposed in (Pedro et al., 2009b), which is designated as Traffic Engineering in the wavelength domain with Delayed Burst Scheduling (TE-DBS), combines the wavelength contention minimization capability of HMPI with selectively delaying data bursts at the electronic buffers of their ingress nodes not only to smooth burst traffic, but also to maximize the amount of data bursts carried in the wavelength channels ranked with the highest priorities by HMPI.

The key principles of the TE-DBS strategy can be illustrated with the example of Fig. 4. The OBS network depicted comprises six nodes and five fibre links. Three paths, π_1, π_2, and π_3,

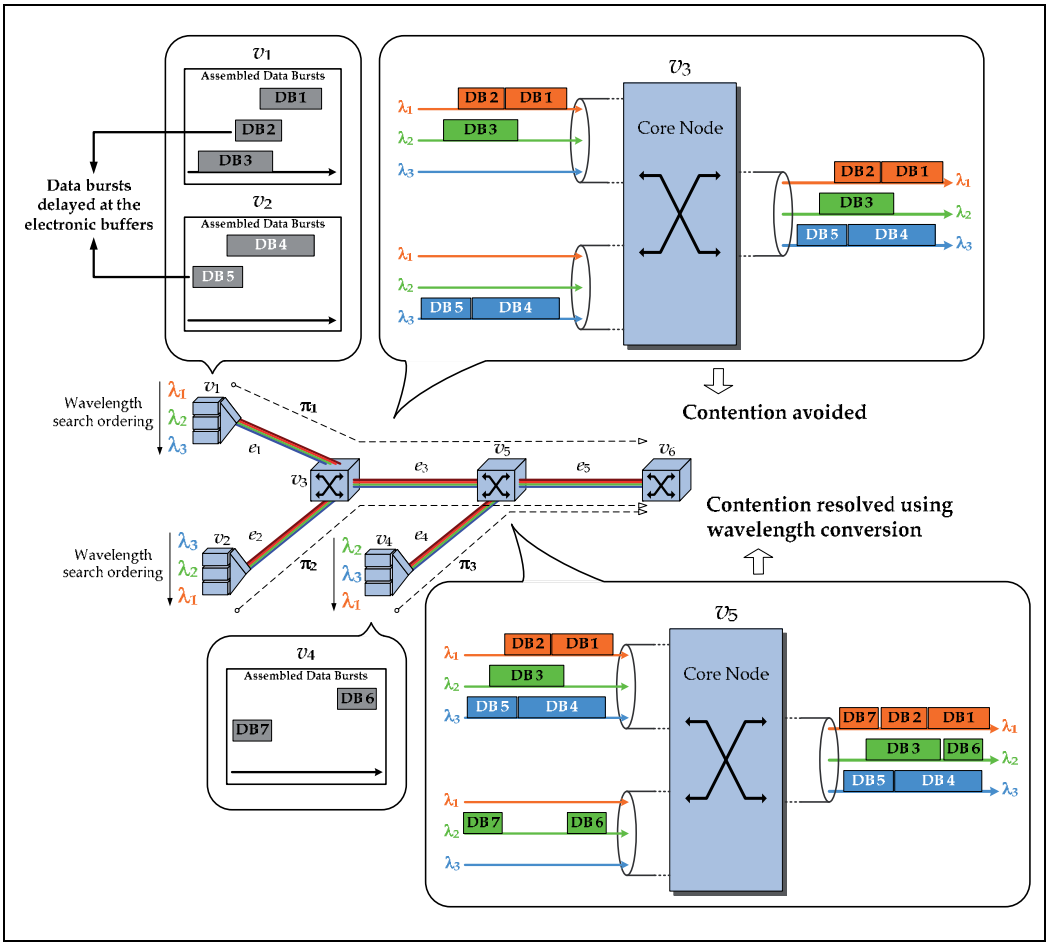


Fig. 4. Example of using TE-DBS to minimize contention at the core nodes.

are used to transmit bursts between one of the three ingress nodes, v_1, v_2, and v_4, and node v_6. Contention between bursts from different input fibre links and directed to the same output fibre link can occur at core nodes v_3 and v_5. Each ingress node uses its own wavelength search ordering and selectively delays bursts with the purpose of transmitting them on the wavelength channels which have been ranked with the highest priorities by an algorithm for minimizing contention in the wavelength domain. Similarly to what occurs with BORA, a maximum ingress burst delay, $\Delta t_{\max}^{\mathrm{RAM}}$, is imposed at each ingress node.

As can be seen, v_1 has assembled three data bursts (DB 1, DB 2, and DB 3), which overlap in time, and v_2 has assembled two data bursts (DB 4 and DB 5), which also overlap in time. The first two bursts assembled by v_1 are transmitted in wavelength channel λ_1, whereas the third cannot be transmitted in this wavelength without infringing the maximum ingress burst delay and, therefore, has to be transmitted in λ_2. The two bursts assembled by v_2 are transmitted in the wavelength ranked with highest priority, λ_3. These bursts traverse v_3, where contention is avoided since the bursts arrive in different wavelengths. Meanwhile, the ingress node v_4 has assembled two data bursts (DB 6 and DB 7) and transmits them in the wavelength ranked with highest priority, λ_2. All seven data bursts traverse core node v_5, where DB 7 must be converted to another wavelength in order to resolve contention.

The major observations provided by this example are as follows. Similarly to using BORA, the burst traffic is smoothed at the ingress nodes, reducing contention at the core nodes from an excessive number of data bursts directed to the same output fibre link. Moreover, since the burst traffic of routing paths π_1, π_2, and π_3 is mostly carried in different wavelengths, contention for the same wavelength channel is also reduced. As a result, the pairs of bursts serialized at the ingress nodes, DB 1 and DB 2 in routing path π_1 and DB 4 and DB5 in routing path π_2, can be kept in the same wavelength channel until they reach node v_6, mitigating the fragmentation of the capacity of wavelengths λ_1 and λ_3 in the fibre links traversed by routing paths π_1 and π_2. Since this is accomplished through minimizing the probability of wavelength contention, it can also relax the wavelength conversion capabilities of the core nodes without significantly degrading network performance.

The TE-DBS strategy requires the computation of one wavelength search ordering, $\{\lambda^1(\pi_i), \lambda^2(\pi_i), \ldots, \lambda^W(\pi_i)\}$, for each routing path π_i. The HMPI algorithm is used to optimize offline the wavelength search orderings. These orderings are stored at the ingress nodes and the control unit of these nodes uses them for serializing data bursts on the available wavelength channel ranked with the highest priority on the routing path the bursts will follow.

4. Results and discussion

This section presents a performance analysis of the framework for traffic engineering in the wavelength domain TE-DBS, described in the Section 3, and assuming the HMPI algorithm, detailed in Section 2, is employed offline to optimize the wavelength search ordering for each routing path in the network.

The results are obtained via network simulation using the event-driven network simulator described in (Pedro et al., 2006a). The network topology used in the performance study is a 10-node ring network. All of the network nodes have the functionalities of both edge and

core nodes and the resource reservation is made using the JET protocol. It is also assumed that all the wavelength channels in a fibre link have a capacity μ = 10 Gb/s, the time required to configure an optical space switch matrix is t_g = 1.6 μs, each node can process the BHP of a data burst in t_p = 1 μs and the offset time between BHP and data burst is given by $t_g + h_i \cdot t_p$, where h_i is the number of hops of burst path $\pi_i \in \Pi$. The switch matrix of each node is assumed to be strictly non-blocking. Unless stated otherwise, the simulation results were obtained assuming W = 32 wavelength channels per fibre link.

The traffic pattern used in the simulations is uniform, in the sense that a burst generated at an ingress node is randomly destined to one of the remaining nodes. Bursts are always routed via the shortest path. Both the data burst size and the burst interarrival time are negative-exponentially distributed. An average burst size of 100 kB is used, which results in an average burst duration of 80 μs. In the network simulations, increasing the average offered traffic load is obtained through reducing the average burst interarrival time. The average offered traffic load normalized to the network capacity is given by,

$$\Gamma = \frac{\sum_{\pi_i \in \Pi} \gamma_i \cdot h_i^{SP}}{L \cdot W \cdot \mu}, \tag{12}$$

where h_i^{SP} is the number of links traversed between the edge nodes of $\pi_i \in \Pi$.

In OBS networks, the most relevant performance metric is the average burst blocking probability, which measures the average fraction of burst traffic that is discarded by the network. The network performance can also be evaluated via the average offered traffic load that results in an objective average burst blocking probability B_{obj}. This metric is estimated by performing simulations with values of Γ spaced by 0.05, determining the load values between which the value with blocking probability B_{obj} is located and then using linear interpolation (with logarithmic scale for the average burst blocking probability). All of the results presented in this section were obtained through running 10 independent simulations for calculating the average value of the performance metric of interest, as well as a 95% confidence interval on this value. However, these confidence intervals were found to be so narrow that have been omitted from the plots for improving readability.

The majority of OBS proposals assumes the utilization of full-range wavelength converters deployed in a dedicated configuration, that is, one full-range wavelength converter is used at each output port of the switch matrix, as illustrated in Fig. 5. Each full-range wavelength converter must be capable of converting any wavelength at its input to a fixed wavelength at its output and if a node has M output fibres, its total number of converters is $M \cdot W$.

Fig. 6 plots the average burst blocking probability as a function of the maximum ingress burst delay for different values of the offered traffic load and considering both TE-DBS and the previously described BORA strategy. It also displays the blocking performance that corresponds to delaying bursts at the ingress nodes whenever a free wavelength channel is not immediately found. More precisely, the DBS strategy consists of delaying a data burst at its ingress node by the minimum amount of time, upper-bounded to the maximum ingress burst delay, such that one wavelength becomes available in the output fibre link.

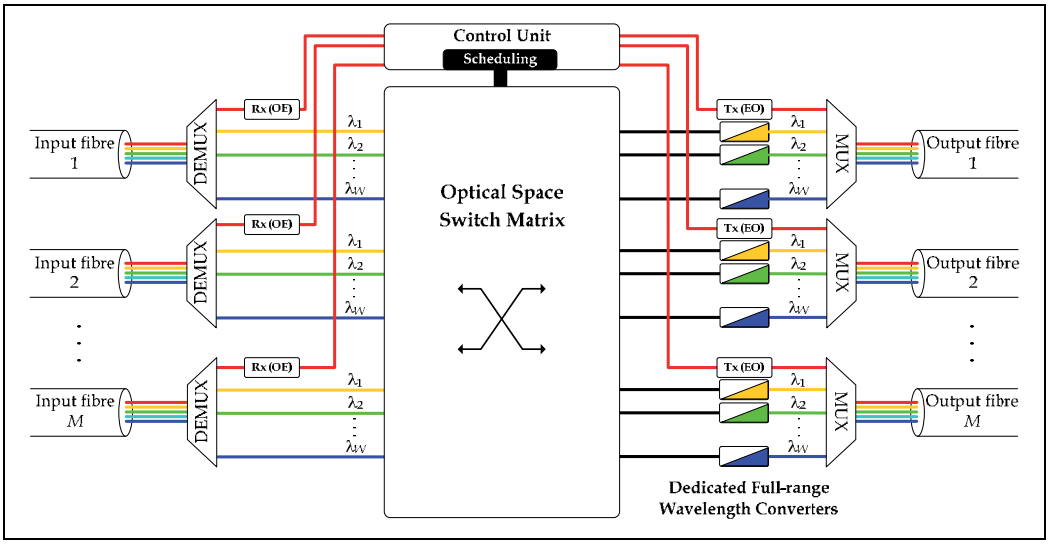


Fig. 5. OBS core node architecture with dedicated full-range wavelength converters.

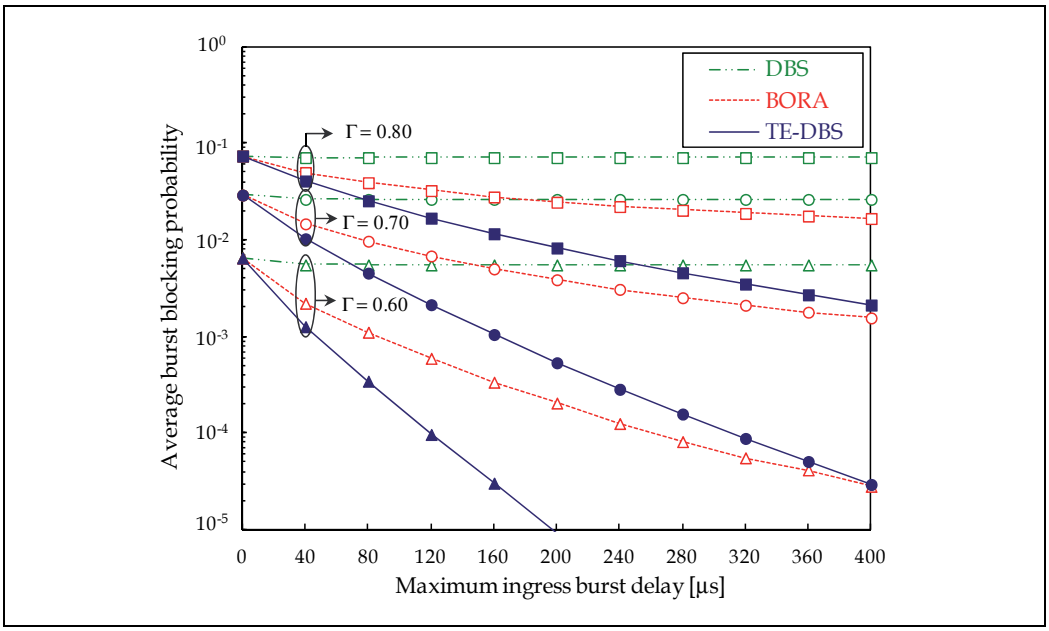


Fig. 6. Network performance with dedicated full-range wavelength converters for different values of the average offered traffic (Pedro et al., 2009a).

The curves for DBS show that exploiting the electronic buffers at the ingress nodes only for contention resolution does not improve blocking performance. On the contrary, with both BORA and TE-DBS the average burst blocking probability is decreased as the maximum ingress burst delay is increased, confirming that these strategies proactively reduce the probability of contention by selectively delaying bursts at their ingress nodes.

The results also indicate TE-DBS is substantially more efficient than BORA in exploiting larger maximum ingress burst delays to reduce the burst blocking probability. The proposed strategy outperforms BORA for the same maximum ingress burst delay or, alternatively, requires a smaller maximum ingress burst delay to attain the same blocking performance of BORA. Particularly, the decrease rate of the burst losses with increasing the maximum ingress burst delay is considerably larger with TE-DBS than that with BORA. In addition, with TE-DBS the slope of the curves of the burst blocking probability is much steeper for smaller values of the average offered traffic load, a trend less pronounced with BORA.

Table 3 presents the average traffic load that can be offered to the network as to support an objective average burst blocking probability, B_{obj}, of 10^{-3} and 10^{-4}. The results include two values of the maximum ingress burst delay for BORA and TE-DBS, Δt_{max}^{RAM} = 200 μs and Δt_{max}^{RAM} = 400 μs, and the case of immediate burst scheduling at the ingress nodes, $\Delta t_{max}^{RAM} = 0$.

B_{obj}	$\Delta t_{max}^{RAM} = 0$	$\Delta t_{max}^{RAM} = 200$ μs		$\Delta t_{max}^{RAM} = 400$ μs	
		BORA	TE-DBS	BORA	TE-DBS
10^{-3}	0.522	0.654	0.723	0.689	0.782
10^{-4}	0.453	0.584	0.659	0.632	0.729

Table 3. Average offered traffic load for an objective average burst blocking probability of 10^{-3} and 10^{-4} (Pedro et al., 2009a).

The OBS network supports more offered traffic load for the same average burst blocking probability when using the TE-DBS and BORA strategies instead of employing immediate burst scheduling. In addition, the former strategy provides the largest improvements in supported offered traffic load. For instance, with $B_{obj} = 10^{-3}$, the network supports 32% more offered traffic load when using BORA with a maximum ingress burst delay of 400 μs instead of immediate burst scheduling, whereas when using the TE-DBS strategy the performance improvement is more expressive, enabling an increase of 50% in offered traffic load.

In order to provide evidence of the principles underlying contention minimization with BORA and TE-DBS, the first set of results differentiates the burst blocking probability at the ingress nodes (ingress bursts) and at the core nodes (transit bursts). Fig. 7 plots the average burst blocking probability, discriminated in terms of ingress bursts and transit bursts, as a function of the maximum ingress burst delay for $\Gamma = 0.70$.

The plot shows that without additional delays at the ingress nodes, the blocking probability of ingress bursts and of transit bursts are of the same order of magnitude. However, as the maximum ingress burst delay is increased, the blocking probability of ingress bursts is rapidly reduced, as a result of the enhanced ability of ingress nodes to buffer bursts during longer periods of time. This holds for the three channel scheduling algorithms. Therefore, the average burst blocking probability of transit bursts becomes the dominant source of blocking. Notably, using DBS does not reduce burst losses at the core nodes, rendering this strategy useless, whereas BORA and TE-DBS strategies exploit the selective ingress delay to reduce blocking of transit bursts. Moreover, TE-DBS is increasingly more effective than BORA in reducing these losses, which supports its superior performance displayed in Fig. 6.

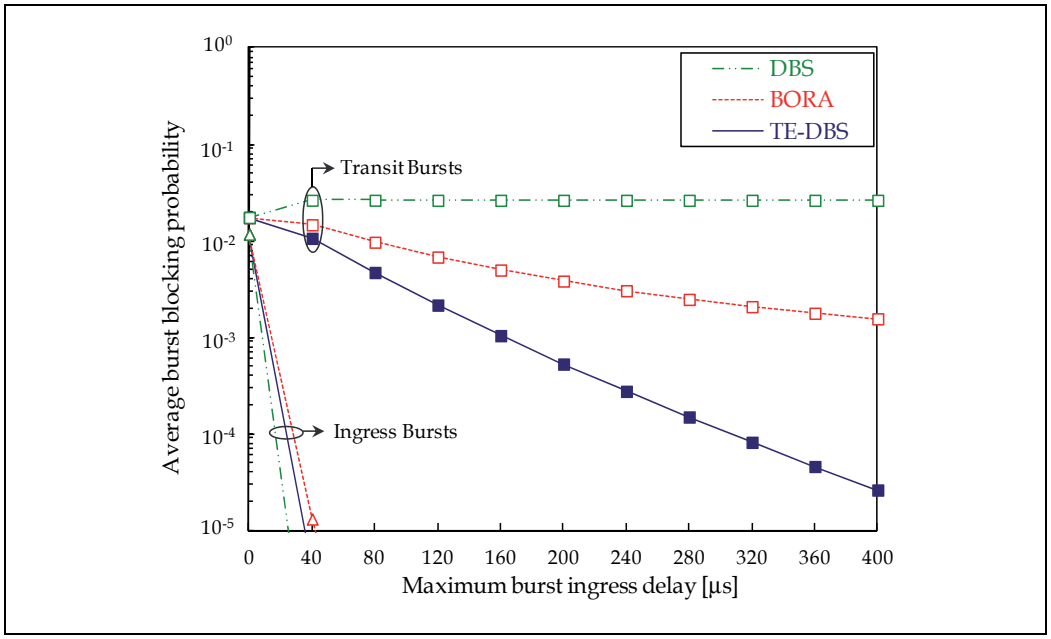


Fig. 7. Average burst blocking probability of ingress and transit bursts (Pedro et al., 2009a).

The major dissimilarity between the TE-DBS and BORA strategies is the order by which free wavelength channels are searched to schedule the data bursts assembled at the ingress nodes. Particularly, the TE-DBS strategy exploits the selective delaying of data bursts at the electronic buffers of these nodes not only to smooth the burst traffic entering the core network, similarly to BORA, but also to proactively reduce the unusable voids formed between consecutive data bursts scheduled in the same wavelength channel. As described in Section 3, complying with the latter objective demands enforcing that the serialized data bursts are kept in the same wavelength for as long as possible along their routing path, which means that contention for the same wavelength among bursts on overlapping paths must be minimized. Intuitively, the success of keeping the serialized data bursts in the same wavelength channel for as long as possible should be visible in the form of a reduced number of bursts experiencing wavelength conversion at the core nodes. In order to observe this effect, Fig. 8 presents the average wavelength conversion probability, defined as the fraction of transit data bursts that undergo wavelength conversion, as a function of the maximum ingress burst delay for different values of the average offered traffic load.

The curves for TE-DBS exhibit a declining trend as the maximum ingress burst delay increases, with this behaviour being more pronounced for smaller average offered traffic load values. These observations confirm that the probability of the data bursts serialized at the ingress nodes being kept in the same wavelength channel, as they go through the core nodes, is higher for larger values of the maximum ingress burst delay and smaller values of offered traffic load. Conversely, with BORA the wavelength conversion probability remains insensitive to variations in both the maximum ingress burst delay and offered traffic load, corroborating the fact that it cannot reduce the utilization of wavelength conversion at the core nodes. The reduced wavelength contention characteristic of the TE-DBS strategy, which

is absent in BORA, is critical to mitigate the fragmentation of the wavelengths capacity, resulting in the smaller transit burst losses reported with TE-DBS in Fig. 7 and ultimately explaining the enhanced blocking performance provided by this strategy.

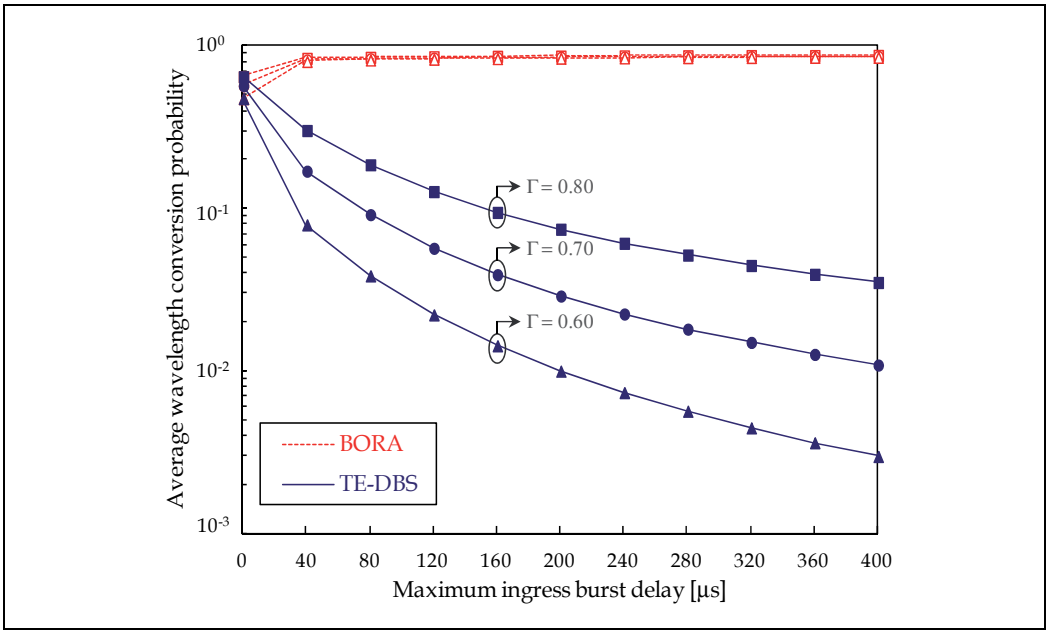


Fig. 8. Average wavelength conversion probability (Pedro et al., 2009b).

Fig. 9 shows the blocking performance as a function of the maximum ingress burst delay for different numbers of wavelength channels and Γ = 0.80. The results indicate that the slope of the average burst blocking probability curves for TE-DBS increases with the number of wavelength channels, augmenting the performance gain of using this strategy instead of BORA. This behaviour is due to the fact that when the number of wavelength channels per fibre link increases the effectiveness of the HMPI algorithm in determining appropriate wavelength search orderings improves, enhancing the isolation degree of serialized burst traffic from overlapping routing paths on different wavelength channels.

In principle, only a fraction of transit bursts experience wavelength contention, demanding the use of a wavelength converter. Consequently, the deployment of a smaller number of converters, in a shared configuration, has been proposed in the literature. Converter sharing at the core nodes can be implemented on a per-link or per-node basis, depending on whether each converter can only be used by bursts directed to a specific output link or can be used by bursts directed to any output link of the node (Chai et al., 2002). The latter sharing strategy enables to deploy a smaller number of converters. Fig. 10 exemplifies the architecture of a core node with C full-range wavelength converters shared per-node, where $C \leq M \cdot W$. In this core node architecture, each wavelength converter must be capable of converting any wavelength channel at its input to any wavelength channel at its output and the switch matrix has to be augmented with C input ports and C output ports.

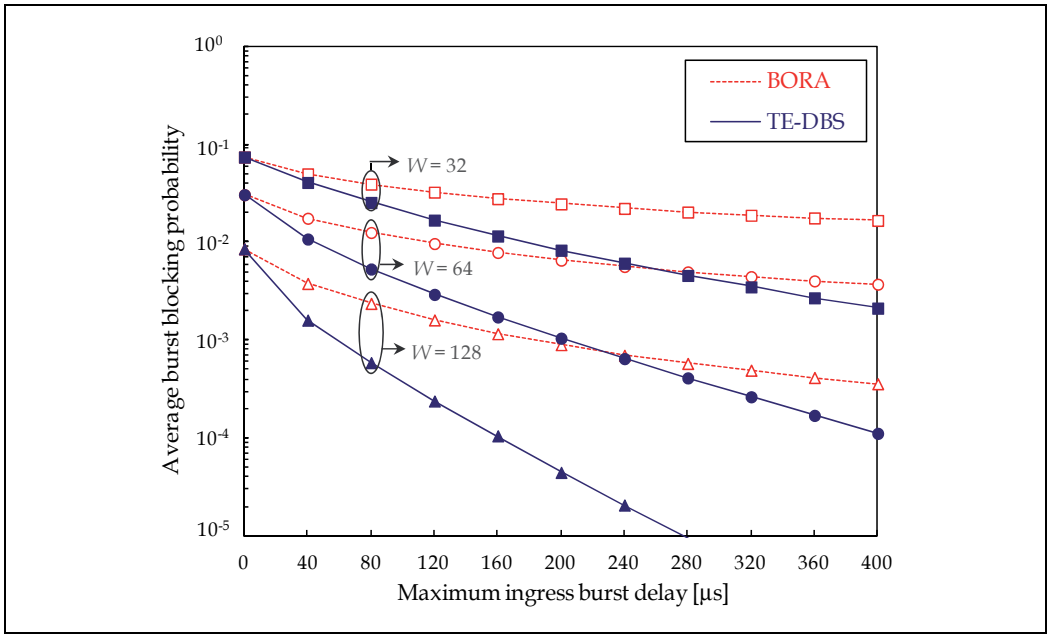


Fig. 9. Network performance for different numbers of wavelength channels (Pedro et al., 2009b).

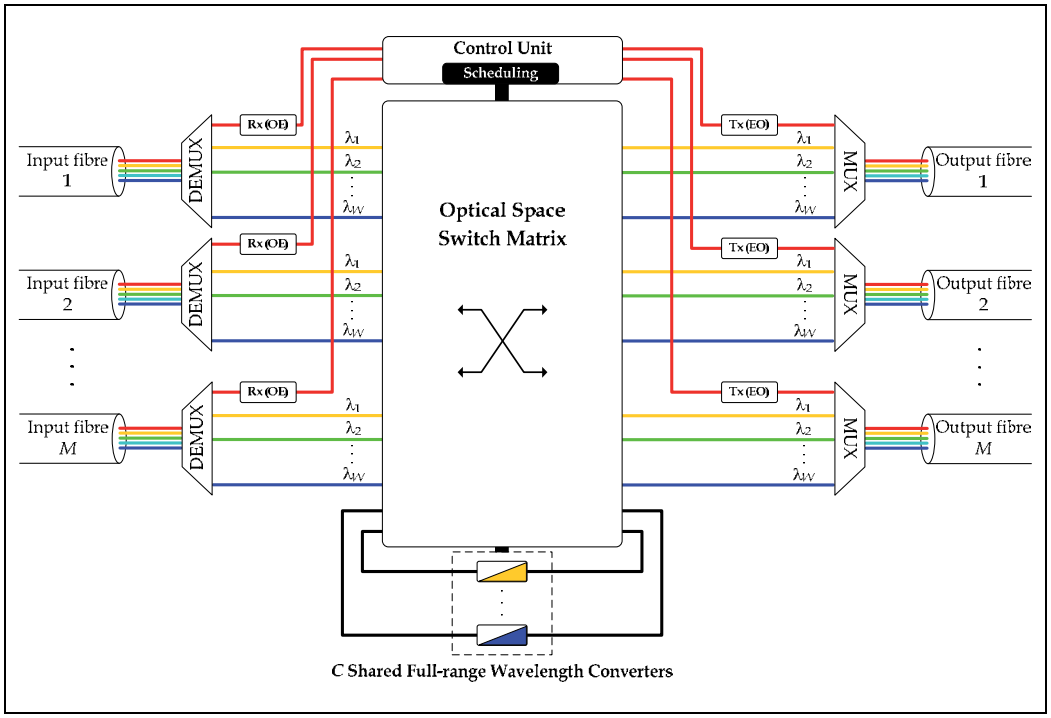


Fig. 10. OBS core node architecture with shared full-range wavelength converters.

The minimization of wavelength contention experienced by transit bursts is a key enabler for TE-DBS to improve the loss performance of OBS networks. Particularly, the simulation results presented in Fig. 8 confirm that the utilization of this strategy reduces the probability of wavelength conversion, and consequently the utilization of the wavelength converters, as the maximum ingress burst delay is increased. This attribute can extend the usefulness of TE-DBS to OBS networks with shared full-range wavelength converters because in this network scenario the lack of available converters at the core nodes can become the major cause of unresolved contention, specially for small values of C.

In order to illustrate the added-value of the TE-DBS strategy in OBS networks whose core nodes have shared full-range wavelength converters, consider the 10-node ring network with $W = 32$. When using wavelength converters in a dedicated configuration, each node of this network needs $M \cdot W = 64$ converters. Fig. 11 plots the average burst blocking probability as a function of the number of shared full-range wavelength converters at the nodes, C, for different values of the average offered traffic load and using BORA and TE-DBS strategies with $\Delta t_{\max}^{\mathrm{RAM}} = 160$ μs.

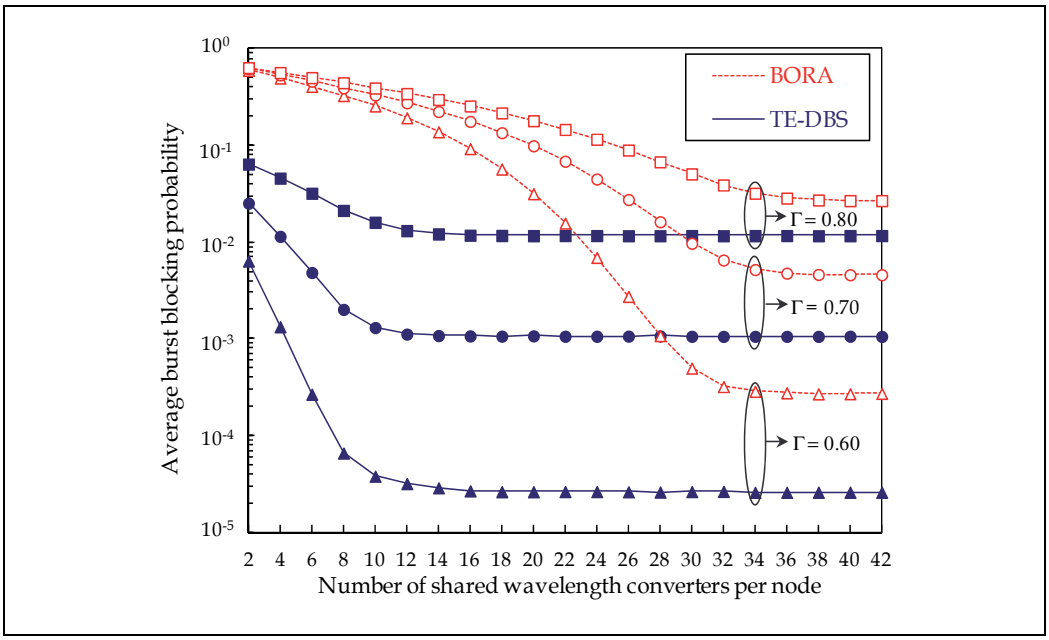


Fig. 11. Network performance with shared full-range wavelength converters for different values of the average offered traffic load (Pedro et al., 2009a).

The blocking performance curves clearly show that the OBS network using TE-DBS can benefit not only in terms of enhanced blocking performance, but also from enabling using simplified core node architectures. More precisely, the burst loss curves indicate that for very small numbers of shared wavelength converters, the utilization of TE-DBS results in a burst blocking probability that can be multiple orders of magnitude lower than that obtained using BORA. Furthermore, using TE-DBS demands a much smaller number of shared wavelength converters to match the blocking performance of a network using core

nodes with dedicated wavelength converters. Particularly, with TE-DBS around 16 shared converters per node are enough to match the loss performance obtained with 64 dedicated converters, whereas with BORA this number more than doubles, since around 36 shared converters are required. The larger savings in the number of wavelength converters enabled by TE-DBS also mean that the expansion of the switch matrix to accommodate the shared converters is smaller, leading to an even more cost-effective network solution.

5. Conclusions

Optical burst switching is seen as a candidate technology for next-generation transport networks. This chapter has described and analyzed the performance benefits of a strategy to enforce traffic engineering in the wavelength domain in OBS networks. The TE-DBS strategy is based on using the HMPI algorithm to optimize offline the order by which wavelength channels are searched for each routing path and employing at the ingress nodes a selective delaying of data bursts as a way to maximize the amount of burst traffic sent via the wavelength channels ranked with highest priority. Both the HMPI offline algorithm and the online selective delaying of bursts were revisited and exemplified.

A network simulation study has highlighted the performance improvements attained by using TE-DBS in an OBS network with dedicated full-range wavelength converters and with shared full-range wavelength converters. It was shown that the utilization of the TE-DBS strategy enables to reduce the average burst blocking probability for a given average offered traffic load, or augment the average offered traffic load for an objective burst blocking probability, when compared to utilizing a known contention minimization strategy. The simulation results shown that increasing the maximum delay a burst can experience at the ingress node and augmenting the number of wavelength channels per link can improve the effectiveness of the TE-DBS strategy and also provided evidence of the burst serialization and traffic isolation in different wavelengths inherent to this strategy. Finally, the analysis confirms that the utilization of TE-DBS in OBS networks with shared full-range wavelength converters can provide noticeable savings in the number of expensive all-optical wavelength converters and a smaller increase in the size of the switch matrix of the core nodes.

6. References

Barakat, N. & Darcie, T. (2007). The Control-Plane Stability Constraint in Optical Burst Switching Networks. *IEEE Communications Letters*, Vol. 11, No. 3, (March 2007), pp. 267-269, ISSN 1089-7798

Chai, T.; Cheng, T. ; Shen, G.; Bose, S. & Lu, C. (2002). Design and Performance of Optical Cross-Connect Architectures with Converter Sharing. *Optical Networks Magazine*, Vol. 3, No. 4, (July/August 2002), pp. 73-84, ISSN 1572-8161

Chang, G.; Yu, J. ; Yeo, Y.; Chowdhury, A. & Jia, Z. (2006). Enabling Technologies for Next-Generation Packet-Switching Networks. *Proceedings of the IEEE*, Vol. 94, No. 5, (May 2006), pp. 892-910, ISSN 0018-9219

Chen, Y.; Qiao, C. & Yu, X. (2004). Optical Burst Switching: A New Area in Optical Networking Research. *IEEE Network*, Vol. 18, No. 3, (May/June 2004), pp. 16-23, ISSN 0890-8044

IETF (2002). *RFC 3945: Generalized Multi-Protocol Label Switching (GMPLS) Architecture*, Internet Engineering Task Force, September 2002

ITU-T (2006). *Recommendation G.8080: Architecture for the Automatically Switched Optical Network (ASON)*, International Telecommunication Union – Telecommunication Standardization Sector, June 2006

Korotky, S. (2004). Network Global Expectation Model: A Statistical Formalism for Quickly Quantifying Network Needs and Costs. *IEEE/OSA Journal of Lightwave Technology*, Vol. 22, No. 3, (March 2004), pp. 703-722, ISSN 0733-8724

Li, J. & Qiao, C. (2004). Schedule Burst Proactively for Optical Burst Switched Networks. *Computer Networks*, Vol. 44, (2004), pp. 617-629, ISSN 1389-1286

Papadimitriou, G.; Papazoglou, C. & Pomportsis, A. (2003). Optical Switching: Switch Fabrics, Techniques, and Architectures. *IEEE/OSA Journal of Lightwave Technology*, Vol. 21, No. 2, (February 2003), pp. 384-405, ISSN 0733-8724

Pedro, J.; Castro, J.; Monteiro, P. & Pires, J. (2006a). On the Modelling and Performance Evaluation of Optical Burst-Switched Networks, *Proceedings of IEEE CAMAD 2006 11th International Workshop on Computer-Aided Modeling, Analysis and Design of Communication Links and Networks*, pp. 30-37, ISBN 0-7803-9536-0, Trento, Italy, June 8-9, 2006

Pedro, J.; Monteiro, P. & Pires, J. (2006b). Wavelength Contention Minimization Strategies for Optical-Burst Switched Networks, *Proceedings of IEEE GLOBECOM 2006 49th Global Telecommunications Conference*, paper OPNp1-5, ISBN 1-4244-0356-1, San Francisco, USA, November 27-December 1, 2006

Pedro, J.; Monteiro, P. & Pires, J. (2009a). On the Benefits of Selectively Delaying Bursts at the Ingress Edge Nodes of an OBS Network, *Proceedings of IFIP ONDM 2009 13th Conference on Optical Network Design and Modelling*, ISBN 978-1-4244-4187-7, Braunschweig, Germany, February 18-20, 2009

Pedro, J.; Monteiro, P. & Pires, J. (2009b). Contention Minimization in Optical Burst-Switched Networks Combining Traffic Engineering in the Wavelength Domain and Delayed Ingress Burst Scheduling. *IET Communications*, Vol. 3, No. 3, (March 2009), pp. 372-380, ISSN 1751-8628

Pedro, J.; Monteiro, P. & Pires, J. (2009c). Traffic Engineering in the Wavelength Domain for Optical Burst-Switched Networks. *IEEE/OSA Journal of Lightwave Technology*, Vol. 27, No. 15, (August 2009), pp. 3075-3091, ISSN 0733-8724

Poustie, A. (2005). Semiconductor Devices for All-Optical Signal Processing, *Proceedings of ECOC 2005 31st European Conference on Optical Communication*, Vol. 3, pp. 475-478, ISBN 0-86341-543-1, Glasgow, Scotland, September 25-29, 2005

Qiao, C. & Yoo, M. (1999). Optical Burst Switching (OBS) – A New Paradigm for an Optical Internet. *Journal of High Speed Networks*, Vol. 8, No. 1, (January 1999), pp. 69-84, ISSN 0926-6801

Sahara, A.; Shimano, K.; Noguchi, K.; Koga, M. & Takigawa, Y. (2003). Demonstration of Optical Burst Data Switching using Photonic MPLS Routers operated by GMPLS Signalling, *Proceedings of OFC 2003 Optical Fiber Communications Conference*, Vol. 1, pp. 220-222, ISBN 1-55752-746-6, Atlanta, USA, March 23-28, 2003

Sun, Y.; Hashiguchi, T. ; Minh, V.; Wang, X.; Morikawa, H. & Aoyama, T. (2005). Design and Implementation of an Optical Burst-Switched Network Testbed. *IEEE*

Communications Magazine, Vol. 43, No. 11, (November 2005), pp. s48-s55, ISSN 0163-6804

Teng, J. & Rouskas, G. (2005). Wavelength Selection in OBS Networks using Traffic Engineering and Priority-Based Concepts. *IEEE Journal on Selected Areas in Communications*, Vol. 23, No. 8, (August 2005), pp. 1658-1669, ISSN 0733-8716

Tucker, R. (2006). The Role of Optics and Electronics in High-Capacity Routers. *IEEE/OSA Journal of Lightwave Technology*, Vol. 24, No. 12, (December 2006), pp. 4655-4673, ISSN 0733-8724

Wang, X.; Morikawa, H. & Aoyama, T. (2003). Priority-Based Wavelength Assignment Algorithm for Burst Switched WDM Optical Networks. *IEICE Transactions on Communications*, Vol. E86-B, No. 5, (2003), pp. 1508-1514, ISSN 1745-1345

Xiong, Y.; Vandenhoute, M. & Cankaya, H. (2000). Control Architecture in Optical Burst-Switched WDM Networks. *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 10, (October 2000), pp. 1838-1851, ISSN 0733-8716

Zang, H.; Jue, J. ; Sahasrabuddhe, L.; Ramamurthy, R. & Mukherjee, B. (2001). Dynamic Lightpath Establishment in Wavelength-Routed WDM Networks. *IEEE Communications Magazine*, Vol. 39, No. 9, (September 2001), pp. 100-108, ISSN 0163-6804

Zhou, P. & Yang, O. (2003). How Practical is Optical Packet Switching in Core Networks?, *Proceedings of IEEE GLOBECOM 2003 49th Global Telecommunications Conference*, pp. 2709-2713, ISBN 0-7803-7974-8, San Francisco, USA, December 1-5, 2003

Zhu, K.; Zhu, H. & Mukherjee, B. (2005). *Traffic Grooming in Optical WDM Mesh Networks,* Springer, ISBN 978-0-387-25432-6, New York, USA

13

Modelling a Network Traffic Probe Over a Multiprocessor Architecture

Luis Zabala, Armando Ferro,
Alberto Pineda and Alejandro Muñoz
University of the Basque Country (UPV/EHU)
Spain

1. Introduction

The need to monitor and analyse data traffic grows with increasing network usage by businesses and domestic users. Disciplines such as security, quality of service analysis, network management, billing and even routing require traffic monitoring and analysis systems with high performance. Thus, the increasing bandwidth in data networks and the amount and variety of network traffic have increased the functional requirements for applications that capture, process or store monitored traffic. Besides, the availability of capture hardware (monitoring cards, taps, etc.) and mass storage solutions at a reasonable cost makes the situation better in the field of network traffic monitoring. For these reasons, several research groups are studying how to monitor heterogeneous network environments, such as wired broadband backbone networks, next generation cellular networks, high-speed access networks or WLAN in campus-like environments. In keeping with this line, our research group NQaS (Networking, Quality and Security) aims to contribute in this challenge and presents theoretical and experimental research to study the behaviour of a probe (Ksensor) that can perform traffic capturing and analysis tasks in Gigabit Ethernet networks. Not only do we intend to progress in the design of traffic analysis systems, but we also want to obtain mathematical models to study the performance of these devices.

The widespread of 1/10 Gigabit Ethernet networks, emphasizes the problems related to system losses which invalidate the results for certain analyses. New Gigabit networks, even at 40 and 100 Gbps, are already being implemented and the problem becomes accentuated. On top of that, commodity systems are not optimized for monitoring [Wang&Liu, 2004] and, as a result, processing resources are often wasted on inefficient tasks. Because of this, new research works have arisen focusing on the development of analysis systems that are able to process all the information carried by actual networks.

Taking all this into account, we would like to develop analytical models that represent traffic monitoring systems in order to provide solutions to the problems mentioned before. Modelling helps to predict the system's performance when it is subjected to a variety of network traffic load conditions. Designers and administrators can identify bottlenecks, deficiencies and key system parameters that impact its performance, and thereby the system can be properly tuned to give the optimal performance. By means of modelling technique, it

is possible to draw qualitative and, in many cases, also quantitative conclusions about features related to modelled systems even without having to develop them. The impact of developing costs, which is a determining factor in some cases, can be dramatically reduced by using modelling.

Having this in mind, and considering the experience of our group, we present our original design (Ksensor) that improves system performance, as well as a mathematical model based on a closed queueing network which represents the behaviour of a multiprocessor traffic monitoring and analysis system. Both things are considered together in the validation of the model, where Ksensor is used as well as a testing platform developed by NQaS. All these aspects are presented throughout this chapter.

A number of papers has addressed the issue of modelling traffic monitoring systems. However, there are more related to the hardware and software involved in this type of systems.

Regarding hardware proposals, one of the most relevant was the development of the high-performance DAG capture cards [Cleary et al., 2000] at the University of Waikato (New Zealand). Several research works and projects have made use of these cards for traffic analysis system design. Some other works proposed the use of Network Processors (NP) [Intel, 2002]. Conventional hardware also showed bottlenecks and new input/output architectures were proposed, such as Intel's CSA (Communication Streaming Architecture).

At the software level, Mogul and Ramakrishnan [Mogul&Ramakrishnan, 1996] identified the most important performance issues on interrupt-driven capture systems. Zero-copy architectures are also remarkable [Zhu et al, 2006]. They try to omit the path followed by packets through the system kernel to the user-level applications, providing a direct access to captured data or mapping memory spaces (mmap). Biswas and Sinha proposed a DMA ring architecture [Biswas&Sinha, 2006] shared by user and kernel levels. Luca Deri suggests a passive traffic monitoring system over general purpose hardware at Gbps speeds (nProbe). Deri has also suggested improvements for the capture subsystem of GNU/Linux, such as a driver-level ring [Deri, 2004], and a user-level library, nCap [Deri, 2005a]. Recently, Deri has proposed a method for speeding up network analysis applications running on Virtual Machines [Cardigliano, 2011], and has presented a framework [Fusco&Deri, 2011] that can be exploited to design and implement this kind of applications.

Other proposals focus on parallel systems. Varenni et al. described the logic architecture of a multiprocessor monitoring system based on a circular capture buffer [Varenni et al.,2003] and designed an SMP driver for DAG cards. We must also remark the KNET module [Lemoine et al., 2003], a packet classifying system at the NIC to provide independent per connection queues for processors. In addition, Schneider and Wallerich studied the performance challenges over general purpose architectures and described a methodology [Schneider, 2007] for evaluating and selecting the ideal hardware/software in order to monitor high-speed networks.

Apart from the different proposals about architectures for capture and analysis systems, there are analytical studies which aim at the performance evaluation of these computer systems. Among them, we want to underline the works done by the group led by Salah

[Salah, 2006][Salah et al., 2007]. They analyse the performance of the capturing system considering CPU consumptions in a model based on queuing theory. Their last contributions explain the evolution of their models towards applications like Snort or PC software routers. Another work in the same line was developed by Wu [Wu et al., 2007], where a mathematical model based on the 'token bucket' algorithm characterized Linux packet reception process.

We also have identified more complex models whose application to traffic capturing and analysis systems can be very beneficial. They are models based on queuing systems with vacations. In this field, we want to underline the contributions from Lee [Lee, 1989], Takagi [Takagi, 1994, 1995] and Fiems [Fiems, 2004].

Most of the previous approaches are for single processor architectures. However, it is clear interest in the construction of analytical models for multiprocessor architectures, in order to evaluate their performance. This paper contributes in this sense from a different point of view, given that the model is based on a closed queueing network. Furthermore, the analytical model and the techniques presented in this paper can be considerably useful not only to model traffic monitoring systems, but also to characterize similarly-behaving queueing systems, particularly those of multiple-stage service. These systems may include intrusion detection systems, network firewalls, routers, etc.

The rest of the chapter is organized as follows: in Section 2 we introduce the framework of our traffic and analysis system called 'Ksensor'. Section 3 presents the analytical model for evaluating the performance of the traffic monitoring system. Section 4 provides details on the analytical solution of the model. Section 5 deals with the validation and obtained results are discussed. Finally, Section 6 remarks the conclusions and future work.

2. Ksensor: Multithreaded kernel-level probe

In a previous work [Muñoz et al., 2007], our research group, NQaS, proposed a design for an architecture able to cope with high-speed traffic monitoring using commodity hardware. This kernel-level framework is called Ksensor and its design is based on the following elements:

- Migration to the kernel which consists in migrating the processing module from user-level to the kernel of the operating system.
- Execution threads defined to take advantage of multiprocessor architectures at kernel-level and solve priority problems. Independent instances are defined for capture and analysis phases. There are as many analysing instances as processors, and as many capturing instances as capturing NICs.
- A single packet queue, shared by all the analysing instances, omitting the filtering module and so saving processing resources for the analysis.

This section explains the main aspects of Ksensor, because of its importance in the validation of the mathematical model which will be explained in a subsequent section.

2.1 Architecture of Ksensor

The kernel-level framework, called Ksensor, intended to exploit the parallelism in QoS algorithms, improving the overall performance.

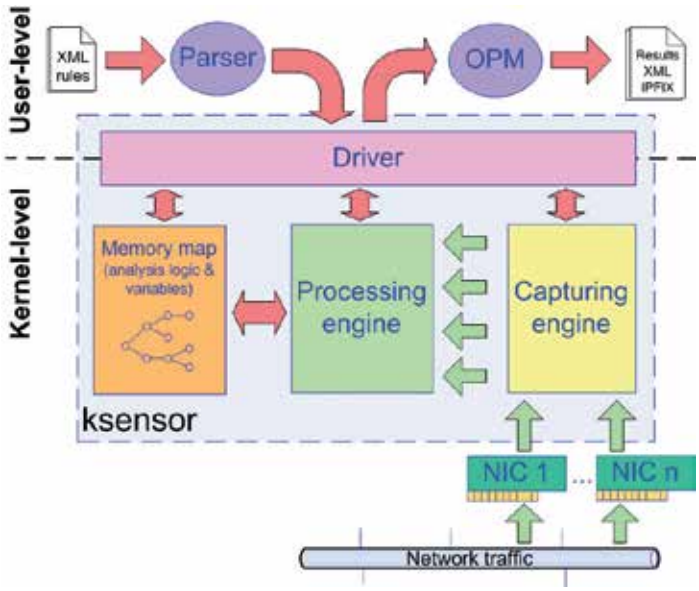


Fig. 1. Architecture of Ksensor.

Fig. 1 shows the architecture of Ksensor. As we can see, only the system configuration (parser) and the result management (Offline Processing Module, OPM) modules are at user-level. Communication between user and Kernel spaces is offered by a module called driver. The figure also shows a module called memory map. This module is shared memory where the analysis logic and some variables are stored.

The definition of execution threads is aimed to take advantage of multiprocessor architectures at kernel-level and solve priority problems, minimizing context and CPU switching. Kernel threads are scheduled at the same level than other processes, so the Kernel's scheduler is responsible for this task.

Ksensor executes two tasks. On one hand, it has to capture network traffic. On the other hand, it has to analyse those captured packets. In order to do that, we define independent instances for capture and analysis phases. Each thread belongs to an execution instance of the system and is always linked with the same processor. All threads share information through the Kernel memory.

In Fig. 2 we can see the multithreaded execution instances in Ksensor. There are as many analysing instances as processors (ksensord#n) and as many capturing instances as capturing NICs (ksoftirqd#n). For example, if the system has two processors, one of them is responsible for capturing packets and analysing some of them and the other one is responsible for analysing packets. This way an analysis task could fill the 100% of one processor's resources if necessary.

The capturing instance takes the packets that the networking subsystem captures and stores them in the packet queue. There is only one packet queue. Processing instances take packets from that queue in order to analyse them.

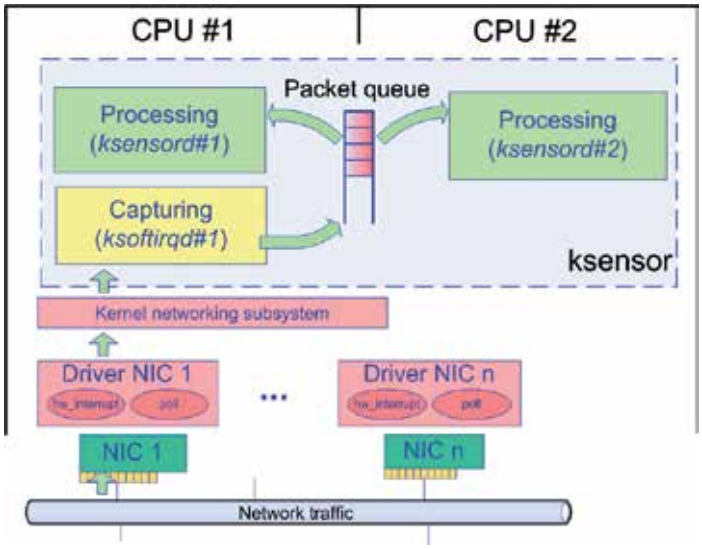


Fig. 2. Multithreaded execution instances in Ksensor.

It does not matter what processing thread analyses a packet because all of them use the same analysis logic. As we said before, there is a shared memory (memory map module) that stores the analysis logic. All the processing threads can access this memory.

2.2 Capturing mechanism in Linux

Ksensor is integrated into the Linux Kernel. In order to capture the packets of the net, Ksensor uses the Kernel networking subsystem. The capturing interface of this subsystem is called NAPI (New API). Nowadays, all the devices have been upgraded to NAPI. Because of that it is important to explain how this interface works [Benvenuti, 2006].

When the first packet arrives to the NIC, it is stored on the card's internal buffer. When the PCI bus is free, the packets are copied from the NIC's buffer to a ring buffer through DMA. The ring buffer is also known as DMA buffer. Once this copy has finished, a hardware interrupt (hardirq) is generated. All of these actions have been executed without consuming any processor's resources.

If the network interface copies a lot of packets in the ring buffer and the Kernel does not take them out, the ring buffer fills up. In this case, unless the interrupts are disabled, another interrupt is generated in order to notify this situation. Then, while the ring buffer is full, the new captured packets will be stored on the NIC's buffer. When this buffer fills up too, the arriving packets will be dropped.

In any case, when the kernel detects the network card interrupt, its handler is executed. In this handler, the NIC driver registers the network interface in an especial list called poll list. This means that this interface has captured packets and needs the Kernel to take them out of the ring buffer. In order to do that, a new software interrupt (softIRQ) is scheduled. Finally, hardIRQs are disabled. From now on, the NIC will not notify new packet arrivals or overload of the ring buffer.

2.3 Network interfaces polling

The softIRQ handler takes out packets from the ring buffer. In Ksensor, after taking out a packet from the ring buffer, the handler stores it in a special queue called packet queue, as we can see in Fig. 2.

The system decides when a softIRQ handler is executed. When its execution starts, the handler polls the first interface in the poll list and starts taking out packets from its ring buffer. In each poll, the softIRQ handler can only pull out packets up to a maximum number called quota. When it reaches the quota it has to poll the next interface in the poll list. If an interface does not have more packets it is deleted from the poll list. Besides, in a softIRQ, the handler can only take out a maximum number of packets called budget. When the handler reaches this maximum, the softIRQ finishes. If there are interfaces left in the poll list, a new softIRQ is scheduled. Furthermore, a softIRQ may take one jiffy (4 ms) at most. If it consumes this time and there are still packets to pull out, the softIRQ finishes and a new one is scheduled.

There is only one poll list in each processor. When the hardIRQ handler is called it registers the network interface in the poll list of the processor that is executing the handler. The softIRQ handler is executed in the same processor. At any given time, a network interface can only be registered in one poll list.

Ksensor has a system to improve the performance in case of congestion. When the packet queue reaches a maximum number of stored packets, this system forces NAPI to stop capturing packets. This means that all the resources of all the processors are dedicated to analysing instances. When the number of packets in the packet queue reaches a fixed threshold value the system starts capturing again.

3. Model for a traffic monitoring system

This section introduces an analytical model which works out some characteristics of network traffic analysis systems. There are several alternatives to model theoretically this type of system. For example, you can use models of queuing theory, Petri nets and, even, mixed models. The ultimate goal is to have a theoretical model that allows us to study the performance of a network traffic analysis system, considering those parameters that are the most representative: throughput, number of processors, analysis load and so on.

We have chosen a theoretical model based on closed queuing networks. It is able to represent accurately the behaviour of a system in charge of analysing network traffic loaded in a multiprocessor architecture. Queuing theory allows us to develop models in order to study the performance of computer's systems [Kobayashi, 1978]. Proposed model consists in a closed queue network where CPU consumptions are related to the service capacity of the queues.

It is worth mentioning that both the flowing traffic and the processing capacity at the nodes are modelled by Poisson arrival rates and exponential service rates. Poisson's distributions are considered to be acceptable for modelling incoming traffic [Barakat et al., 2002]. This assumption can be relaxed to more general processes such as MAPs (Markov Arrival Processes) [Altman et al., 2000], or non homogeneous Poisson processes, but we will keep working with it for simplicity of the analysis. Regarding service rate modelling, although

program's code has a quite deterministic behaviour, some randomness is introduced by Poisson incoming traffic, variable length of packets and kernel scheduler uncertainty.

3.1 Description of the model

The proposed queuing network for modelling a traffic monitoring system is showed in Fig. 3. It consists of two parts; the upper one has a set of multi-server queues which represents the processing ability of the traffic analysis system. The lower part models the injection of network traffic with λ rate with a simple queue. The number of packets that are permitted in the closed queue network is fixed and its value is N.

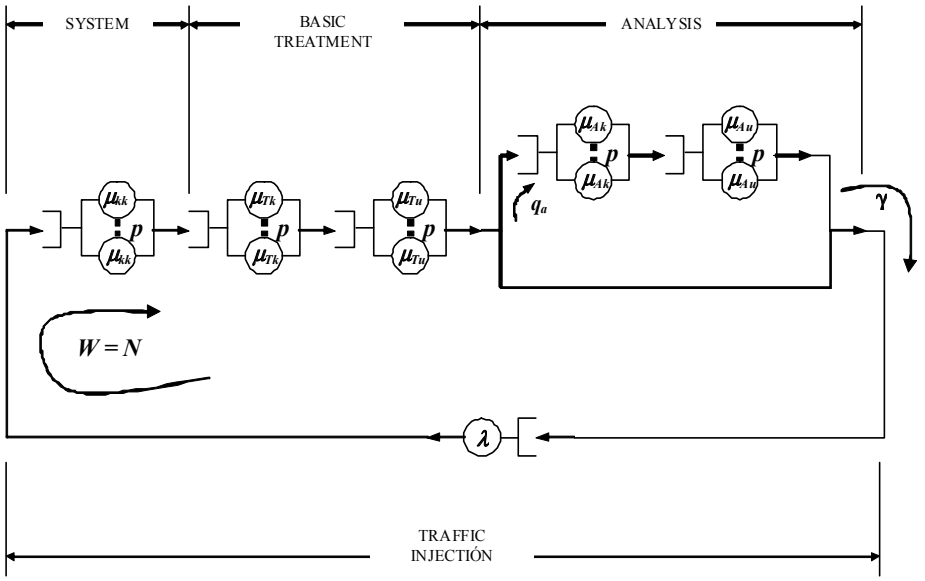


Fig. 3. General model for the traffic analysis system.

Some stages are divided into multiple queues, due to the need to differentiate the processing done in the Kernel and the processing done at user level. Although the process code is usually running on the user level, system calls that require Kernel services are also used.

Four different stages have been distinguished for the closed network, each one with a specific function:

- System stage (system queue): it consists in a queue of μ_{kk} (measured in packets per second) capacity. This stage represents the time spent on the Kernel level of the operating system by the traffic analysis system. It comprises treatments of device controllers and attention paid by kernel to interruptions (hardIRQ and softIRQ) due to packet arrival.
- Basic treatment stage (treatment queues): it is modelled by two queues with μ_{Tk} and μ_{Tu} capacities. This stage represents the amount of time consumed by the system to perform basic treatment to packets captured from the net. This is mainly accomplished by studying control headers of the packets and by determining through a decision tree whether a packet need to be further analysed or not.

- Analysis stage (analysis queues): it is integrated by two queues with μ_{Ak} and μ_{Au}. This stage simulates the analysis treatment that the system does to packets that need further analysis. Not all the packets need to be analysed in this stage. For this reason, a rate called q_a has been defined to represent the proportion of received packet that has to be analysed.
- Traffic injection stage (injection queue): it is a simple queue of λ capacity. This stage simulates the arrival of packets to the system with a λ rate. Since the number of packets in the closed network is fixed to N, the traffic injection queue can be empty. This situation simulates the blocking and new packets will not be introduced on the system.

Each service queue has p servers that represent the p processors of a multiprocessor system. Multiple server representation has been chosen to emphasize the possibility of parallelizing every stage of processing. However, all stages may not be necessarily parallelizable. For example, only one processor can access NIC at the same time, so the packet capturing process will not be parallelizable in different instances.

Another aspect to consider is that packets cannot flow freely in the closed network, because the sum of packets attended in the servers that represent the traffic monitoring system never exceeds the maximum number of processors available. Therefore, we have to assure that, at any time, the maximum number of packets in the upper queues of Fig 3 is not greater than p (the number of processors).

Considering an arrival rate of λ packets per second, the traffic analysis system will be able to keep pace with a part of that traffic, defined as $q \cdot \lambda$. Remaining traffic ($(1-q) \cdot \lambda$) will be lost because the platform is not capable of dealing with all the packets. Captured traffic, $q \cdot \lambda$, goes through the system and basic treatment stages. Nevertheless, all traffic will not be subject of further analysis because of features of the modelled system. For example, a system in charge of calculating QoS parameters of all connections that arrive to a server will discard the packets with other destination address or monitoring systems which use sampling techniques will discard a percentage of packets or intrusion detection will apply further detection techniques only to suspicious packets. Therefore, q_a coefficient has been defined to represent the rate of captured packets liable of being further analysed (analysis stage) than treated only (treatment stage). Thus, $q_a \cdot q \cdot \lambda$ of the initial flow will go through the analysis stage.

3.2 Simplifications of the model

The model presented in Fig. 3 is very general, but if we observe it, some simplifications are possible. Simplifications allow us to group different service rates to identify parameters that may be analysed easily. Among the possible simplifications, we highlight two: one related to CPU consumption and another one, to the equivalent traffic monitoring system.

3.2.1 Model of CPU consumption

This simplification proposes to group all the kernel consumptions in a simple queue, whereas user processes consumptions are represented in a multi-queue. It considers that kernel services are hardly parallelizable.

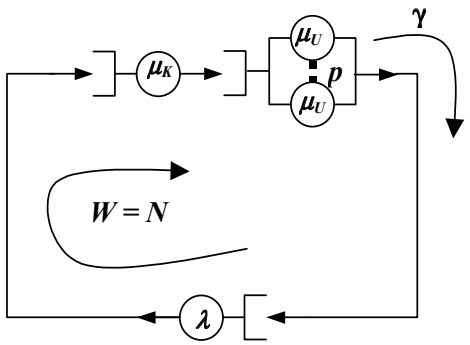


Fig. 4. Model of CPU consumption.

The equivalent service rates can be calculated as follows.

$$\frac{1}{\mu_K} = \frac{1}{\mu_{pk}} + \frac{1}{\mu_{Kk}} = \frac{q_a}{\mu_{Ak}} + \frac{1}{\mu_{Tk}} + \frac{1}{\mu_{Kk}} \tag{1}$$

$$\frac{1}{\mu_U} = \frac{1}{\mu_{pu}} = \frac{q_a}{\mu_{Au}} + \frac{1}{\mu_{Tu}} \tag{2}$$

3.2.2 Model of the equivalent traffic monitoring system

The main feasible simplification preserving the identity of the system is to replace the whole system with an equivalent multi-server queue applying the Norton equivalence [Chandy et al., 1975]. The Norton theorem establishes that in networks with solution in product form, any subnetwork can be replaced by a queue with a state-dependent service capacity. Our theoretical model has exponential service rates in all stages, so applying the Norton equivalence, the new equivalent queue will have a state-dependent service capacity $\mu_{eq}(n,q_a)$.

The simple queue μ_S of the Fig. 5 represents non-parallelizable processes of the system and the multiple queue μ_M represents parallelizable ones.

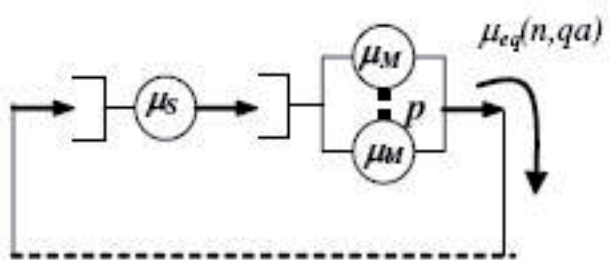


Fig. 5. Traffic monitoring system that Norton equivalence is applied to.

This model adapts perfectly to Ksensor, because we identify a non-parallelizable process that corresponds with the packet capture and parallelizable processes that are related to analysis. Both μ_S and μ_M (in packets per second) can be measured in the laboratory.

4. Analytical study of the model

This section presents the analytical study of the model. It can be directly addressed by analytical calculation, assuming Poisson arrivals and exponential service times. Perhaps the greatest difficulty lies in determining the abstractions that are necessary to adapt the model to the actual characteristics of the traffic monitoring system. Likewise, we propose a method of calculation based on mean value analysis which allows us to solve systems with more elements, where the analytical solution may be more complex to develop.

4.1 Equations of the general model

Viewing the simplifications that have been developed, we might observe that, in the study of this model, a topology is repeated at different levels of abstraction. This topology corresponds with a closed network model with two queues in series; first, a simple one, and second, another one with multiple servers, as shown in Fig. 6. This structure usually occurs in every processing stage. Processing at Kernel level is usually not parallelizable, and therefore, the model is represented as a simple queue. On the other hand, the user processing is usually parallelizable and it is represented by a multiple queue with p servers, being p the number of processor of the platform. The appearance of this topology allows us to define a simple model that we can solve analytically.

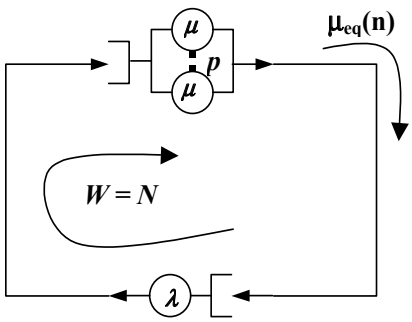


Fig. 6. Closed queue network simplified for the general model.

In order to get the total throughput of the system, first, we calculate the state probabilities for the network, putting N packets in circulation through the closed network, but assuming that the upper multiple queue can have at most p packets being served and the rest waiting in the queue. We also assume that the service capacity in every state of the multiple queue is not proportional to the number of packets. Thus, we will consider μ_i as the service capacity for the state i. The state diagram for this topology is presented in Fig. 7. In this model we are representing the state i of the multiple queue. N packets are flowing through the closed network and we refer to the state i when there are i packets in the multiple queue and the rest, N-i, in the simple queue. The probability of that state is represented as p_i. Finally, the simple queue with rate λ is the packet injection queue.

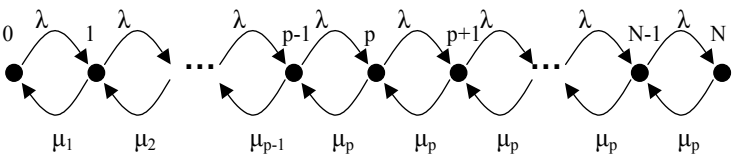


Fig. 7. State diagram for the multiple queue.

It is possible to deduce the balance equations from the diagram of states and, subsequently, the expression of the probability of any state i as a function of the probability of zero state p_0:

$$\forall i = 1,\cdots,p \Rightarrow \left\{ \begin{array}{l} p_0 \cdot \lambda = p_1 \cdot \mu_1 \\ p_1 \cdot \lambda = p_2 \cdot \mu_2 \\ \cdots\cdots\cdots\cdots \\ p_{p-1} \cdot \lambda = p_p \cdot \mu_p \end{array} \right\} \Rightarrow p_i = \frac{\lambda}{\mu_i} \cdot p_{i-1} \tag{3}$$

$$\Rightarrow p_i = \overbrace{\frac{\lambda}{\mu_i} \cdot \frac{\lambda}{\mu_{i-1}} \cdots \frac{\lambda}{\mu_1}}^{\text{i terms}} \cdot p_0 = \frac{\lambda^i}{\prod_{j=1}^{i} \mu_j} \cdot p_0 \tag{4}$$

From this equation, we deduce p_p, the probability of the state p:

$$\Rightarrow p_p = \frac{\lambda^p}{\prod_{j=1}^{p} \mu_j} \cdot p_0 \tag{5}$$

For the states with i>p, their probabilities can be expressed as:

$$\forall i = p+1,\cdots,N \Rightarrow \left\{ \begin{array}{l} p_p \cdot \lambda = p_{p+1} \cdot \mu_p \\ p_{p+1} \cdot \lambda = p_{p+2} \cdot \mu_p \\ \cdots\cdots\cdots\cdots\cdots \\ p_{N-1} \cdot \lambda = p_N \cdot \mu_p \end{array} \right\} \Rightarrow p_i = p_{i-1} \cdot \frac{\lambda}{\mu_p} \tag{6}$$

$$p_i = \overbrace{\frac{\lambda}{\mu_p} \cdot \frac{\lambda}{\mu_p} \cdots \frac{\lambda}{\mu_p}}^{\text{(i-p) terms}} \cdot p_p = \left(\frac{\lambda}{\mu_p} \right)^{i-p} \cdot p_P \tag{7}$$

From this equation we can also derive the expression of the probability p_N, which is interesting because it indicates the probability of having all the packets in the multiple queue and there is none in the simple queue. This probability defines the blocking probability (P_B) of the simple queue.

$$p_N = P_B = \frac{\lambda^N}{\mu_p^{N-p} \cdot \prod_{j=1}^{p} \mu_j} \cdot p_0 \tag{8}$$

Applying the normalization condition (the sum of all probabilities must be equal to 1), we can obtain the general expression for p_0 and, then, we get every state probabilities.

$$\sum_{i=0}^{N} p_i = 1 = p_0 + \sum_{i=1}^{p} p_i + \sum_{i=p+1}^{N} p_i \tag{9}$$

$$1 = p_0 + p_0 \sum_{i=1}^{p} \frac{\lambda^i}{\prod_{j=1}^{i} \mu_j} + p_0 \frac{\lambda^p}{\prod_{j=1}^{p} \mu_j} \cdot \sum_{i=p+1}^{N} \frac{\lambda^{i-p}}{\mu_p^{i-p}} \tag{10}$$

$$\Rightarrow p_0 = \left(1 + \sum_{i=1}^{p} \frac{\lambda^i}{\prod_{j=1}^{i} \mu_j} + \frac{\lambda^p}{\prod_{j=1}^{p} \mu_j} \cdot \sum_{i=p+1}^{N} \frac{\lambda^{i-p}}{\mu_p^{i-p}} \right)^{-1} \tag{11}$$

Considering equations (8) and (11), we have the following blocking probability p_N.

$$p_N = \frac{\lambda^N \Big/ \mu_p^{N-p}}{\left(\prod_{j=1}^{p} \mu_j + \sum_{i=1}^{p} \left(\lambda^i \cdot \prod_{j=i}^{p} \mu_j \right) + \lambda^p \cdot \sum_{i=p+1}^{N} \frac{\lambda^{i-p}}{\mu_p^{i-p}} \right)} \tag{12}$$

P_N is the probability of having N packets in the multiple queue (traffic analysis system queue) of Fig. 6 , so there is not any packet in the injection queue. This situation describes the loss of the system. In order to calculate the throughput γ of the system, (13) is used.

$$\gamma = \lambda \cdot (1 - P_N) \tag{13}$$

Taking into account these expressions, which are valid for the general case, we can develop the equations of the model for some particular cases that will be detailed below: the calculation of the equivalence for the traffic monitoring system and the solution for the closed network with incoming traffic load.

4.2 Calculation of the equivalence for the traffic monitoring system

In general, multiprocessor platforms that implement traffic monitoring systems have certain limitations to parallelize some parts of the processing they do. In particular, Kernel services are not usually parallelizable. This means that, despite having a multiprocessor architecture with p processors that can work in parallel, some services will be performed sequentially and we will lose some of the potential of the platform. For all this, in order to calculate the

Norton equivalence for a traffic monitoring system, one must begin with a model that contains a simple queue and a multi-server queue. This is a particular case of the general model studied before.

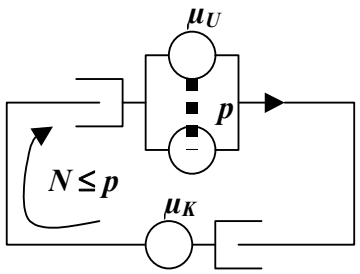


Fig. 8. Equivalence for the traffic monitoring system.

The simple queue with service rate μ_K models non-parallelizable Kernel services, whereas the multiple queue with p servers and service rate μ_U models the system capacity to parallelize certain services. The particularity of this model with regard to the general model is that, at most, only p packets can circulate on the closed network maximum. We are interested in solving this model to work out the equivalent service rate of the traffic monitoring system for every state in the network.

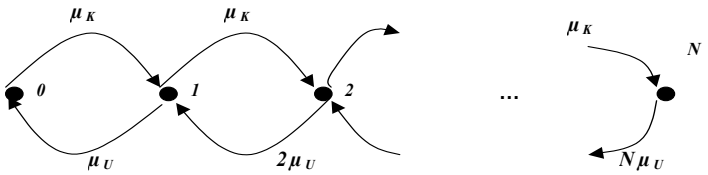


Fig. 9. State diagram for the traffic monitoring system equivalence.

The state diagram makes sense for values of N that are less or equal to the highest number of processors. The service rate of the traffic monitoring system will be different for every value of N and, given that some services are not parallelizable, in general, it does not follow a linear evolution. Following a similar approach to the general case, we can calculate the probability of the highest state, p_N, which is useful to estimate the effective service rate of the equivalence.

$$\left.\begin{array}{c} p_0 \cdot \mu_K = p_1 \cdot \mu_U \\ p_1 \cdot \mu_K = p_2 \cdot 2\mu_U \\ \cdots \\ p_{i-1} \cdot \mu_K = p_i \cdot i \cdot \mu_U \end{array}\right\} \Rightarrow p_i = \frac{\mu_K}{i \cdot \mu_U} \cdot p_{i-1} \tag{14}$$

$$p_i = \frac{\mu_K}{i \cdot \mu_U} \cdot p_{i-1} = \frac{\mu_K^2}{\mu_U^2 \cdot i \cdot (i-1)} \cdot p_{i-2} = \cdots = \frac{\mu_K^i}{\mu_U^i \cdot i!} \cdot p_0 \tag{15}$$

After considering the normalization condition, we can determine the expression for p_N:

$$p_0 + \sum_{i=1}^{N} p_i = 1 = p_0 + \sum_{i=1}^{N} \frac{\mu_K^i}{\mu_U^i \cdot i!} \cdot p_0 = p_0 \cdot \left(1 + \sum_{i=1}^{N} \frac{\rho^i}{i!}\right) \tag{16}$$

$$\Rightarrow p_0 = \frac{1}{\left(1 + \sum_{i=1}^{N} \frac{\rho^i}{i!}\right)} \tag{17}$$

$$p_N = \frac{\mu_K^N}{\mu_U^N \cdot N!} \cdot \frac{1}{\left(1 + \sum_{i=1}^{N} \frac{\rho^i}{i!}\right)} = \frac{\rho^N}{N! + \sum_{i=1}^{N} \frac{N! \cdot \rho^i}{i!}} = \frac{\rho^N}{\sum_{i=0}^{N} \frac{N! \cdot \rho^i}{i!}} \tag{18}$$

Thus, taking into account that the throughput of the closed network is the equivalent service rate, we have the following expression:

$$\mu_{eq}(n) = \mu_K \cdot (1 - p_n) \tag{19}$$

$$\mu_{eq}(n) = \mu_K \cdot \left(1 - \frac{\rho^n}{\sum_{i=0}^{n} \frac{n! \cdot \rho^i}{i!}}\right) \quad // \quad \rho = \frac{\mu_K}{\mu_U} \tag{20}$$

Note that this case is really a particular case of the general case where $\lambda = \mu_K$ and $\mu_i = i \cdot \mu_U$.

4.3 Solution for the closed network model with incoming traffic

The previously explained Norton equivalence takes into consideration the internal problems of the traffic monitoring system related to the non-parallelizable tasks. Now we will complete the model adding the traffic injection queue to the equivalent system calculated before.

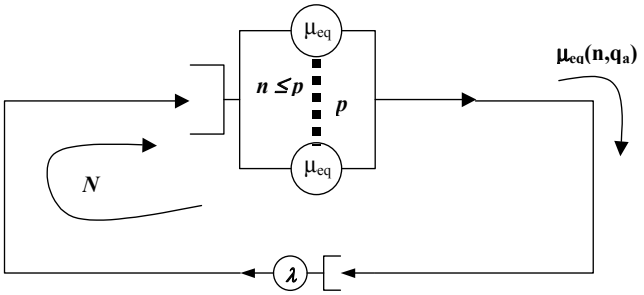


Fig. 10. General model with incoming traffic.

The entire system under traffic load is modelled as a closed network with an upper multiple queue, which is the Norton equivalent queue of the traffic analysis system, and a lower simple queue, simulating the injection of network traffic with rate λ. In this closed network,

a finite number N of packets circulate. In general, this number N is greater than p, the number of available processors.

The analytical solution of this model is similar to that proposed for the general model taking into account that the service rates μ_1, μ_2..., μ_p will correspond with the calculation of the Norton equivalent model $\mu_{eq}(n, q_a)$ with values of n from 1 to p. This model allows us to calculate the theoretical throughput of the traffic monitoring system for different loads of network traffic.

$$\gamma = \lambda \cdot (1 - p_N) \tag{21}$$

The value of N will allow us to estimate the system losses. There will be losses when the N packets of the closed network are located in the upper queue. At that time, the traffic injection queue will be empty and, therefore, it will simulate the blocking of the incoming traffic. That will be less likely, the higher the value of N is.

4.4 Mean value analysis

Apart from the analytic solution explained above, we have also considered an iterative method based on the mean value analysis (MVA), in order to simplify the calculations even more. This theorem states that 'when one customer in an N-customer closed system arrives at a service facility he/she observes the rest of the system to be in the equilibrium state for a system with N−1 customers' [Reiser&Lavengerg, 1980]. The application of this theorem to our case requires taking into account the dependencies between some states and others in a complex state diagram, where the state transitions can be also performed with different probabilities, because there are state dependent service rates.

4.4.1 Probability flows between adjacent states

The mean value analysis is based on the iterative dependency between the probability of a certain state with regard to the probabilities of the closest states. The state transitions will not be possible between any two states, they can only occur between adjacent states.

$$p(i,j) = f\left(p(i-1,j), p(i,j-1)\right) \tag{22}$$

It is necessary to do a balance of probability flows between states considering the service rates that are dependent on the state of each queue.

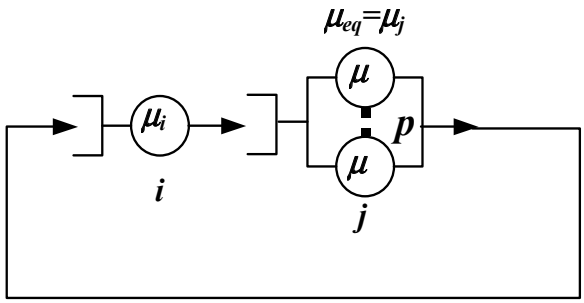


Fig. 11. General model for the closed queue network.

To begin with, we consider the general model for the closed queue network. We call queue i to the simple queue of the model. We assume that this simple queue is in state i and its service rate is μ_i. Likewise, we call queue j to the multi-server queue which is in state j with a state dependent equivalent service rate μ_j. A fixed number of packets (N) are circulating in the closed network, so that there is a dependence between the state i and j.

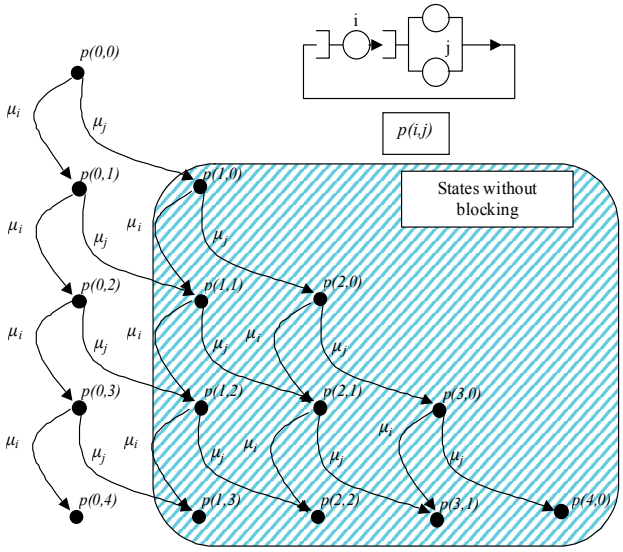


Fig. 12. Probability flows between adjacent states with two processors.

Fig.12 shows the dependencies of the probability of a given state with regard to the closer states in the previous stage with one packet less.

4.4.2 Iterative calculation method

Little's law [Little, 1961] can help us to interpret the relationship between the state probabilities at different stages of the closed queue network.

$$E(T) = \frac{E(n)}{\gamma} \tag{23}$$

This formula is applied to any queue system that is in equilibrium in which there are users who enter the system, consume time to be attended and then leave. In the formula, γ can be understood as the throughput of the system, E(T) as the average time spent in the system and E (n) as the average number of users.

The iterative method applied to the closed queue network is based on solving certain interesting statistics of the network at every stage, using the data obtained in the previous stage. You go from one stage with N packets to the next with N+1 packets, adding one packet to the closed queue network once the system is in stable condition. Knowing the state probability distribution in stage N, we can calculate the average number of users on each server.

$$E(n_i) = \sum_{i=1}^{N} i \cdot p_N(i, N-i) \qquad E(n_j) = \sum_{j=1}^{N} j \cdot p_N(N-j, j) \tag{24}$$

We can calculate every state probability in the stage N as the ratio of the average stay time in this state, $t_N(i,j)$ and the total time for that stage T_{TN}. The total time T_{TN} can be calculated as the sum of all the partial times $t_N(i, j)$ of each state at that stage.

$$p_N(i,j) = \frac{t_N(i,j)}{T_{TOTAL,N}} \tag{25}$$

$$T_{TOTAL,N} = \sum_{i=0}^{N} t_N(i, N-i) \tag{26}$$

If we consider Reiser's theorem [Reiser, 1981], it is possible to set a relation between the state probabilities of a certain state with regard to the ones which are adjacent in the previous stage. In particular, in equilibrium, when we have N packets, the state probability distribution is equal to the distribution at the moment of a new packet arrival at the closed network. In the state diagram of our model, in general, every state depends on two states of the previous stage. We will have the following probability flows:

Transition (i-1,j) → (i,j) a new packet arrives at queue i

$$p'_N(i,j) = p_{N-1}(i-1, j) \tag{27}$$

Transition (i,j-1) → (i,j) a new packet arrives at queue j

$$p''_N(i,j) = p_{N-1}(i, j-1) \tag{28}$$

Knowing the iterative relations of the probabilities between different stages and basing on Little's formula, we can calculate the average stay time $t_N(i, j)$ in the system in a given state, accumulating the average time in queue i, $t^i_n(i, j)$ and the average time in queue j, $t^j_n(i, j)$.

$$t_N(i,j) = t^i_N(i,j) + t^j_N(i,j) \tag{29}$$

Applying Little's law:

$$t^i_N(i,j) = \frac{E^i_N(i)}{\mu_i(i)} = \frac{p'_N(i,j) \cdot i}{\mu_i(i)} = \frac{p_{N-1}(i-1,j) \cdot i}{\mu_i(i)} \tag{30}$$

$$t^j_N(i,j) = \frac{E^j_N(j)}{\mu_j(j)} = \frac{p''_N(i,j) \cdot j}{\mu_j(j)} = \frac{p_{N-1}(i,j-1) \cdot j}{\mu_j(j)} \tag{31}$$

Considering the probability distribution of the previous stage:

$$t_N(i,j) = \frac{p_{N-1}(i-1,j) \cdot i}{\mu_i(i)} + \frac{p_{N-1}(i,j-1) \cdot j}{\mu_j(j)} \tag{32}$$

Taking into account that, for a given state (i, j), the average stay time of a packet in the queues i and j is given by t_i and t_j respectively, we can express the probability of that state as:

$$\tau_i = \frac{i}{\mu_i(i)} \qquad \tau_j = \frac{j}{\mu_j(j)} \tag{33}$$

$$t_N(i,j) = \frac{p_{N-1}(i-1,j)\cdot i}{\mu_i(i)} + \frac{p_{N-1}(i,j-1)\cdot j}{\mu_j(j)} \tag{34}$$

$$p_N(i,j) = \frac{t_N(i,j)}{T_{TN}} = p_{N-1}(i-1,j)\cdot\frac{\tau_i}{T_{TN}} + p_{N-1}(i,j-1)\cdot\frac{\tau_j}{T_{TN}} \tag{35}$$

Eq. 35 allows us to calculate a certain state probability of the stage with N packets, having the probabilities of the adjacent states in the stage N. Using this equation, we can iteratively calculate the state probability distribution for every stage.

4.4.3 Adjusting losses depending on N

The losses of the traffic monitoring system can be measured assessing the blocking probability of the injection queue. If we consider the general model with an incoming traffic of λ, we can calculate (Eq. 21) the volume of traffic processed by the traffic monitoring system (γ) and also the caused losses (δ).

$$\gamma = \lambda\cdot\left(1 - p(0,N)\right) \tag{36}$$

$$\delta = \lambda - \gamma = \lambda\cdot p(0,N) \tag{37}$$

If we look at the evolution of the blocking probability of the injection queue with increasing number of packets N in the closed network, we can see how that probability stage is reduced in each stage. The same conclusion can be derived from Eq. 18.

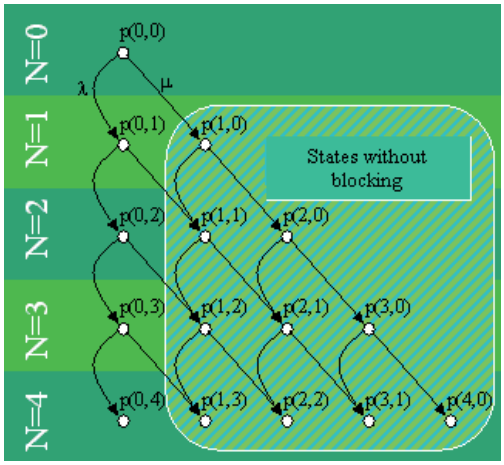


Fig. 13. Evolution of probability flows as a function of N.

A parameter that can be difficult to assess is N, the number of packets that are circulating in the closed network. In general, this parameter depends on specific features of the platform, such as the number of available processors and the ability of the Kernel to accept packets in transit regardless of whether they have processors available at that time.

One conclusion to be drawn from the model, is that it is possible to estimate the value of the parameter N by adjusting the losses that the model has with regard to those which actually occur in a traffic monitoring system.

5. Model validation

This section presents the validation tests to verify the correctness of our analytical model. The aim is to compare theoretical results with those obtained by direct measurement in a real traffic monitoring system, in particular, in the Ksensor prototype developed by NQaS which is integrated into a testing architecture. It is also worth mentioning that, prior to obtaining the theoretical performance results, it is necessary to introduce some input parameters for the model. These initial necessary values will also be extracted from experimental measurements in Ksensor and the testing platform, making use of an appropriate methodology. With all this, we report experimental and analysis results of the traffic monitoring system in terms of two key measures, which are the mean throughput and the CPU utilization. These measures are plotted against incoming packet arrival rate. Finally, we discuss the results obtained.

5.1 Test setup

In this section, we describe the hardware and software setup that we use for our evaluation. Our hardware setup (see Fig. 14) consists of four computers: one for traffic generation (injector), a second one for capturing and analysing the traffic (sensor or Ksensor), a third one for packet reception (receiver) and the last one for managing, configuring and launching the tests (manager). All they are physically connected to the same Gigabit Ethernet switch.

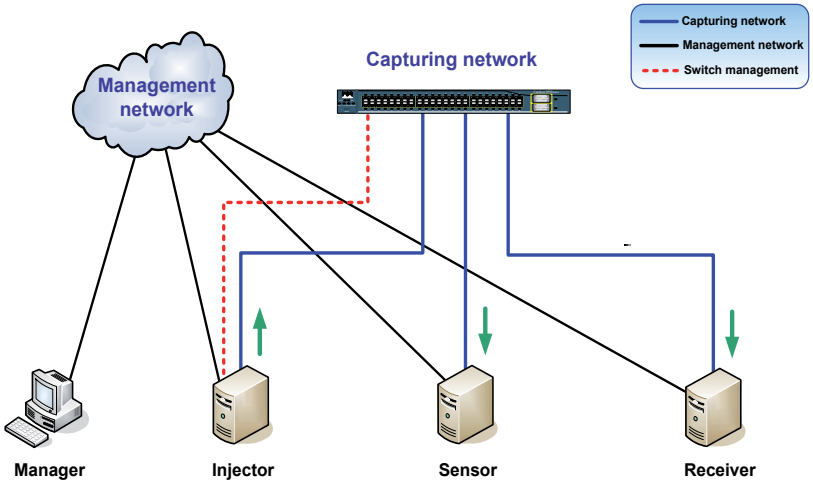


Fig. 14. Hardware setup for validation tests.

However, two virtual networks are distinguished: the first one is the capturing network that connects the elements that play some role during the tests; the second one is the management network which contains the elements that are responsible for the management tasks that can be needed before or after doing tests. The use of two separate networks is necessary, so that the information exchange between the management elements does not interfere with the test results.

The basic idea is to overwhelm Ksensor (sensor) with high traffic generated from the injector. Despite the fact that we do not have 10 Gigabit Ethernet hardware for our tests available, we can achieve our goal of studying the behaviour of the traffic capturing and analysis software at high rates. In addition, we can compare the results with the analytical model and also identify the possible bottlenecks of all analysed systems.

Regarding software, we use a testing architecture [Beaumont et al., 2005] designed by NQaS that allows the automation of tasks like configuration, running and gathering results related to validation tests. The manager, the injector and the sensor that appear in Fig. 14 are part of this testing architecture. They have installed the necessary software to perform the functions of manager, agent, daemon or formatter as we will explain in the next subsection. On the other hand, the receiver is simply the destination of the traffic entered into the network by the injector and it does not have any other purpose.

5.2 Architecture to automatically test a traffic capturing and analysis system

As mentioned previously, in this section, we use a testing architecture for experimental performance measures and, also, to estimate the values of certain input parameters required for the analytical model. It is, therefore, advisable to explain, albeit briefly, the main elements of this platform.

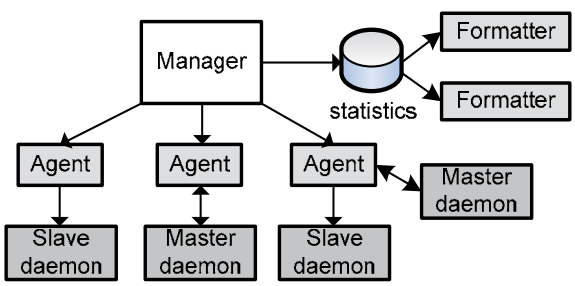


Fig. 15. Logical elements of the testing architecture used in validation tests.

The testing architecture consists of four types of logical elements as Fig. 15 shows. Each of them implements a perfectly defined function:

- Manager is the interface with the user. This element, in the infrastructure shown in Fig. 14, is located on the machine with the same name. It is in charge of managing the rest of the logical elements (agents, daemons and formatters) according to the configuration received from the administrator. After introducing the test setup, it is distributed from the manager to the other elements and the test is launched when the manager sends the start command. At the end of every test, the manager receives and stores the results obtained by the rest of the elements.

- Agents are responsible for attending manager's requests and acting on different devices. Agents are always listening and they have to start and stop the daemons, as well as to collect the performance results. During a test in the infrastructure, one agent is executed in the injector and another one, in the sensor.
- Daemons are in charge of acting on the different physical elements which are involved in each test. Its function can be very variable. For example, the injection of network traffic according to the desired parameterization, the configuration of the capturing buffers, the execution of control programs in the sensor, the acquisition of information or some element's statics, etc. Depending on the relationship with the agent two different types of daemons can be distinguished: master and slave. Master daemons have got some intelligence. The agent will start them but they will indicate when their work has finished. On the other hand, slave daemons do not determine the end of its execution. In each test, to do all the tasks, as many daemons as necessary are executed in the injector and in the sensor.
- Formatters are the programs which select and translate the information stored by the manager to more appropriate formats for its representation. They are executed in the machine called manager, at the end of every test.

5.3 Experimental estimation for certain parameters of the model

In section 3, we have defined an analytical model which functionally responds to a traffic monitoring system. In order to perform an assessment of the model, first we need some values for certain input parameters. We are referring to some service rates that appear in the model based on closed queue networks and are necessary to obtain theoretical performance results. Then we can compare these analytical results with those obtained in the laboratory.

In general, we talk about μ service rates, but, in this subsection, it is easier to talk about mean service times. For this reason, we use the nomenclature based on average processing time in which an average time t_{ij} can be expressed as the inverse of its service rate $1/\mu_{ij}$.

We want to adapt the theoretical model to Ksensor, a real network traffic probe. The best approach is to consider the model of the equivalent traffic monitoring system (see Fig.5) where we distinguish a non-parallelizable process and a parallelizable one. In Ksensor, this separation corresponds with the packet capturing process and the analysis process.

The packet capturing process is not parallelizable because the softIRQ is responsible for the capture and it only runs in one CPU. Fig. 16 shows experimental measurements about average packet capturing times. They have been obtained running tests with Ksensor under different conditions: variable packet injection rate in packets per second and traffic analysis load in number of cycles (null, 1K, 5K or 25K). The inverse of the average softIRQ times shown in Fig. 16 will be the service rate μ_s that appears in the model.

On the other hand, the analysis process is parallelizable in Ksensor. In the same way that softIRQ times have been obtained, we experimentally get average analysis processing times that are shown in Fig. 17. The inverse of the average times shown in Fig.17 will be the service rate μ_M that appears in the multi-queue of the model. It is necessary to comment that, in Fig. 16, the average softIRQ times are not constant. This is because neither all the injected packets are captured by the system, nor all the captured packets are analysed and this causes different computational flow balances.

The values μ_s and μ_M, derived from these experimental measurements, will be taken to the performance evaluation of the model that will be explained later. In addition to the two parameters mentioned, there is another one which is q_a, but it is always q_a=1 in our test configuration.

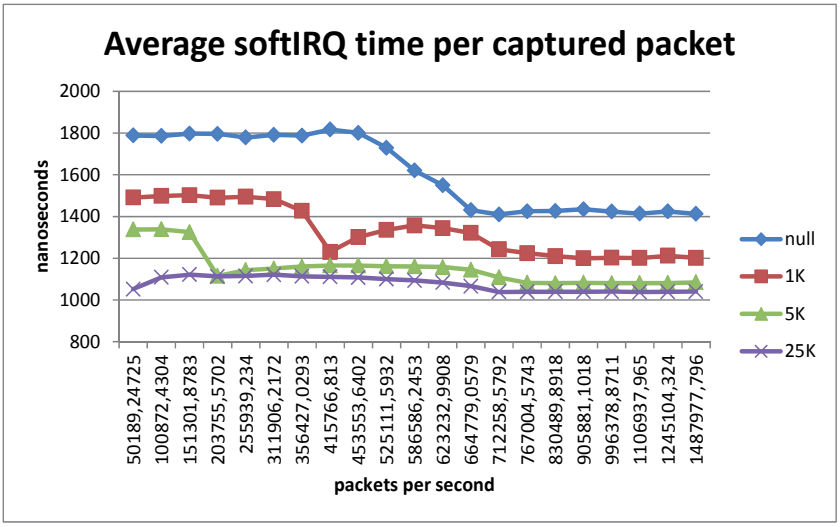


Fig. 16. Average softIRQ per captured packet.

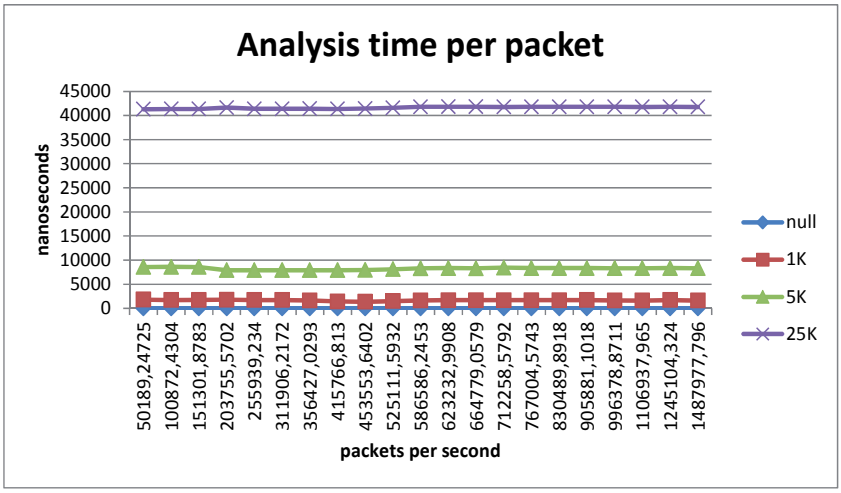


Fig. 17. Analysis time per packet.

5.4 Performance measurements - Evaluation and discussion

The analytical model has been tested with Ksensor under different conditions: packet injection rate (packets per second) varies between 0 and 1.5 million, packet length is 64-1500 bytes and traffic analysis load (at present we simulate QoS algorithm processing times, from 0 to 25000 cycles). The number of processors has been 2 in every test.

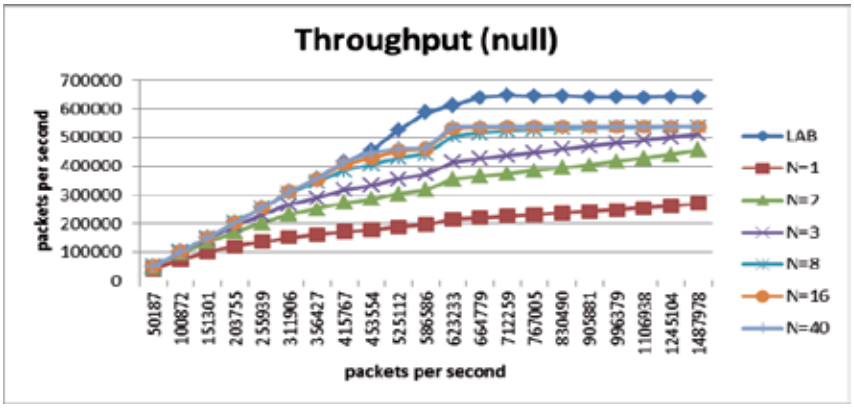


Fig. 18. Theoretical and experimental throughputs without analysis load.

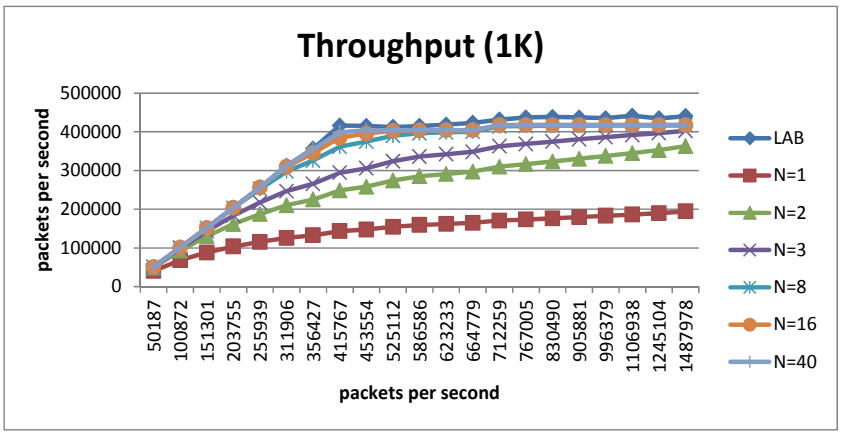


Fig. 19. Theoretical and experimental throughputs with 1Kcycle analysis load.

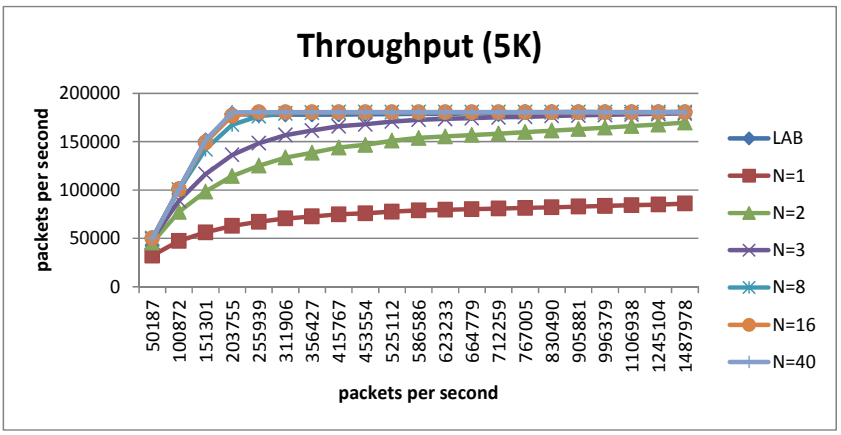


Fig. 20. Theoretical and experimental throughputs with 5Kcycle analysis load.

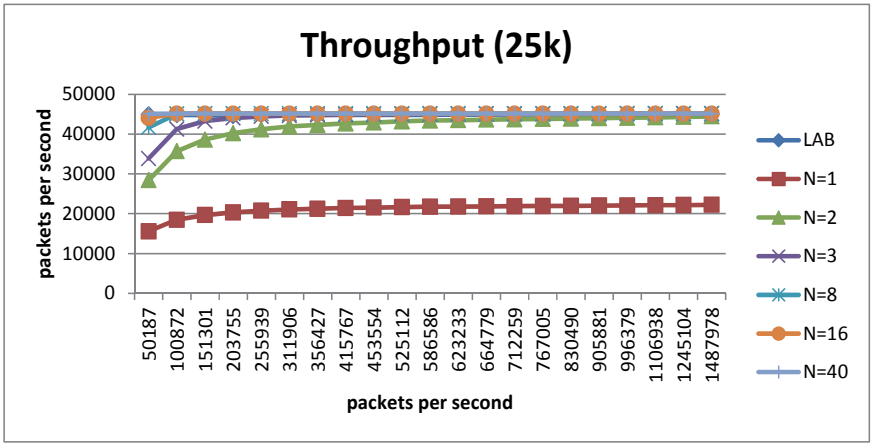


Fig. 21. Theoretical and experimental throughputs with 25Kcycle analysis load.

Fig. 18, Fig. 19, Fig. 20 and Fig. 21 show the comparison between the theoretical model's throughput for different values of N and the real probe's throughput measured experimentally (marked as LAB in the graph). 64 byte-length packets have been used in the lab test and its corresponding service rates in the theoretical calculation. The service rates has been calculated according to the method explained in subsection 5.3.

In all the cases, the throughput grows until a maximum is reached (saturation point). We also observe in these graphs that, with increasing N, the theoretical throughput is close to the real one. It shows, therefore, that the analytical model fits the real system.

6. Conclusion

In this chapter we have presented an analytical model that represents a multiprocessor traffic monitoring system. This model analyses and quantifies the system performance and it can be useful to improve aspects related to hardware and software design. Even, the model can be extended to more complex cases which have not been treated in the laboratory.

Thus, the major contribution of this chapter is the development of a theoretical model based on a closed queuing network that allows to study the behaviour of a multiprocessor network probe. A series of simplifications and adaptations is proposed for the closed network, in order to fit it better to the real system. We obtain the model's analytic solution and we also propose a recurrent calculation method based on the mean value analysis. The model has been validated comparing theoretical results with experimental measures. In the validation process we have made use of a testing architecture that not only has measured the performance, it has also provided values for some necessary input parameters of the mathematical model. Moreover, the architecture helps to setup tests faster as well as to collect and plot results easier. Ksensor, a real probe, is part of the testing architecture and, therefore, it is directly involved in the validation process. As has been seen in the validation section, Ksensor's throughput is acceptably calculated by the model proposed in this chapter. The conclusions obtained have been satisfactory with regard to the behaviour of the model.

This paper has also come in useful to explain the main aspects of Ksensor, a multithreaded kernel-level probe developed by NQaS research group. It is remarkable that this system introduces performance improving design proposals into traffic analysis systems for passive QoS monitoring.

As a future work, we suggest two main lines: the first one is related to Ksensor and it is about a new hardware-centered approach whose objective is to embed our proposals onto programmable network devices like FPGAs. The second research line aims at completing and adapting the model to the real system in a more accurate way. We are already making progress on new mathematical scenarios which can represent, in detail, aspects such as packet capturing process, congestion avoidance mechanisms between capturing and analysis stages, specific analysis algorithms applied in QoS monitoring and packet filtering.

Finally, it is worth mentioning that the test setup, which has been used to validate the model, will be improved acquiring network hardware at 10 Gbps and installing Ksensor over a server with more than two processors. The model will be tested under these new conditions and we hope to obtain satisfactory results, too.

Thus, further work is necessary to analyse this type of systems with a higher precision, compare their results, in certain conditions, better and prevent us from developing high-cost prototypes.

7. References

Altman, E.; Avratchenkov,K. & Barakat, C.. (2000). A stochastic model for TCP/IP with stationary random losses. *ACM SIGCOMM 2000*.

Barakat, C.; Thiran, P.; Iannaccone, G.; Diot, C. & Owezarski, P. (2002). A flow-based model for Internet backbone traffic, *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement, 2002*.

Beaumont, A.; Fajardo, J.; Ibarrola, E. & Perfecto, C. (2005). Arquitectura de red para la automatización de pruebas. *VI Jornadas de Ingeniería Telemática.,* Vigo, Spain.

Benvenuti, C. (2006). *Understanding Linux Network Internals,* O' Reilly Media.

Biswas, A.; Sinha, P. (2006). Efficient real-time Linux interface for PCI devices: A study on hardening a Network Intrusion Detection System. *SANE 2006,* Delft, The Netherlands.

Cardigliano, A. (2011). Towards wire-speed network monitoring using Virtual Machines. *Master Thesis,* University of Pisa, Italy.

Chandy, K.M.; Herzog, U. & Woo, L.S. (1975). Parametric Analysis of Queueing Networks Learning Techniques, *IBM J. Research and Development*, vol. 19, no. 1, pp. 43-49, January 1975.

Cleary, J.; Donnelly, S.; Graham, I.; McGregor, A. & Pearson, M. (2000). Design principles for accurate passive measurement. *Passive and Active Measurement. PAM 2000,* Hamilton, New Zealand.

Deri, L. (2004). Improving Passive Packet Capture: Beyond Device Polling. *SANE 2004,* Amsterdam, The Netherlands.

Deri, L. (2005). nCap: Wire-speed Packet Capture and Transmission. *E2EMON 2005,* Nice, France.

Fiems, D. (2004). Analysis of discrete-time queueing systems with vacations. *PhD Thesis,* Ghent University, Belgium.

Fusco, F. & Deri, L. (2010). High Speed Network Traffic Analysis with Commodity Multi-core Systems. *Internet Measurement Conference 2010,* Melbourne, Australia.

Intel-CSA. (2002). Communication Streaming Architecture: Reducing the PCI Network Bottleneck.

Kobayashi, H. (1978). *Modeling and Analysis: An Introduction to System Performance Evaluation Methodology,* Ed. Wiley-Interscience, ISBN 0-201-14457-3.

Lee, T. (1989). M M/G/1/N queue with vacation time and limited service discipline. *Performance Evaluation,* vol. 9, no. 3, pp. 181-190.

Lemoine, E.; Pham, C. & Lefèvre, L. (2003). Packet classification in the NIC for improved SMP-based Internet servers. *ICN'04,* Guadeloupe, French Caribbean.

Little, J. D. C. (1961). A proof of the queueing formula: L=λ·W, *Operations Research*, vol. 9, no. 3, pp. 383-386, 1961.

Mogul, J.C. & Ramakrishnan, K.K. (1996). Eliminating Receive Livelock in an Interrupt-driven Kernel. *USENIX 1996 Annual Technical Conference,* San Diego, California.

Muñoz, A.; Ferro, A.; Liberal, F. & López, J. (2007). A Kernel-Level Monitor over Multiprocessor Architectures for High-Performance Network Analysis with Commodity Hardware. *SensorComm 2007,* Valencia, Spain.

Reiser, M. (1981). Mean value analysis and convolution method for queue-dependent servers in closed queueing networks, *Performance Evaluation*, vol. 1, no. 1, pp. 7-18, January 1981.

Reiser, M. & Lavengerg, S.S. (1980). Mean Value Analysis of Closed Multichain Queueing Networks, *Journal of the ACM*, vol. 27, no. 2, pp. 313-322, April 1980.

Salah, K. (2006). Two analytical models for evaluating performance of Gigabit Ethernet hosts with finite buffer. *AEU - International Journal of Electronics and Communications,* vol. 60, no. 8, pp. 545-556.

Salah, K.; El-Badawi, K. & Haidari, F. (2007). Performance analysis and comparison of interrupt-handling schemes in gigabit networks. *Computer Communications,* vol. 30, no. 17, pp. 3425-3441.

Schneider, F. (2007). Packet Capturing with Contemporary Hardware in 10 Gigabit Ethernet Environments. *Passive and Active Measurement. PAM 2007,* Louvain-la-Neuve, Belgium.

Takagi, H. (1991). *Queueing Analysis, A Foundation of Performance Evaluation Volume 1: Vacation and Priority Systems (Part 1),* North-Holland, Amsterdam, The Netherlands.

Takagi, H. (1994). M/G/1/N Queues with Server Vacations and Exhaustive Service. *Operations Research,* pp. 926-939.

Varenni, G.; Baldi, M.; Degioanni, L. & Risso, F. (2003). Optimizing Packet Capture on Symmetric Multiprocessing Machines. *15th Symposium on Computer Architecture and High Performance Computing,* Sao Paulo, Brazil.

Wang, P. & Liu, Z. (2004). Operating system support for high performance networking, a survey. *The Journal of China Universities of Posts and Telecommunications,* vol. 11, no. 3, pp. 32-42.

Wu, W; Crawford, M. & Bowden, M. (2007). The performance analysis of linux networking - Packet receiving. *Computer Communications,* vol. 30, no. 5, pp. 1044-1057.

Zhu, H.; Liu, T.; Zhou, C. & Chang, G. (2006). Research and Implementation of Zero-Copy Technology Based on Device Driver in Linux. *IMSCCS'06*.

14

Routing and Traffic Engineering in Dynamic Packet-Oriented Networks

Mihael Mohorčič and Aleš Švigelj
Jožef Stefan Institute
Slovenia

1. Introduction

Spurred by the vision of seamless connectivity anywhere and anytime, ubiquitous and pervasive communications are playing increasingly important role in our daily lives. New types of applications are also affecting behaviour of users and changing their habits, essentially reinforcing the need for being always connected. This clearly represents a challenge for the telecommunications community especially for operating scenarios characterised by high dynamics of the network requiring appropriate routing and traffic engineering.

Routing and traffic engineering are cornerstones of every future telecommunication system, thus, this chapter is concerned with an adaptive routing and traffic engineering in highly dynamic packet-oriented networks such as mobile ad hoc networks, mobile sensor networks or non-geostationary satellite communication systems with intersatellite links (ISL). The first two cases are recently particularly popular for smaller scale computer or data networks, where scarce energy resources represent the main optimisation parameter both for traffic engineering and routing. However, they require a significantly different approach, typically based on clustering, which exceeds the scope of this chapter. The third case, on the other hand, is particularly interesting from the aspect of routing and traffic engineering in large scale telecommunication networks. Even more so, since it exhibits a high degree of regularity, predictability and periodicity. It combines different segments of communication network and generally requires distinction between different types of traffic. Different restrictions and requirements in different segments typically require separate optimization of resource management.

So, in order to explain all routing functions and different techniques used for traffic engineering in highly dynamic networks we use as an example the ISL network, characterized by highly dynamic conditions. Nonetheless, wherever possible the discussion is intentionally kept independent of the type of underlying network or particular communication protocols and mechanisms (e.g. IP, RIP, OSPF, MPLS, IntServ, DiffServ, etc.), although some presented techniques are an integral part of those protocols. Thus, this chapter is focusing on general routing and traffic engineering techniques that are suitable for the provision of QoS in packet-oriented ISL networks. Furthermore, most concepts,

described techniques, procedures and algorithms, even if explained on an example of ISL network, can be generalised and used also in other types of networks exhibiting high level of dynamics (Liu et al., 2011; Long et al., 2010; Rao & Wang, 2010, 2011). The modular approach allows easy (re)usage of presented procedures and techniques, thus, only particular or entire procedures can be used.

ISL network exhibits several useful properties which support the development of routing procedures. These properties include (Wood et al., 2001):

- Predictability – motion of satellites around the earth is deterministic, thus the position of satellites and their connectivity can be computed in advance, taking into account the parameters of the satellite orbit and constellation. Consequently, in an ISL network only undeterministic parameters need to be monitored and distributed through the network, thus minimizing the signalling load.
- Periodicity – satellite positions and thus the configuration of the space segment, repeats with the orbit period, which is defined uniquely by the selected orbit altitude. Taking into account also the terrestrial segment, an ISL network will experience a quasi-periodic behaviour on a larger scale, defined as the smallest common integer multiple of the orbit period and the traffic intensity period, referred to as the system period.
- Regularity – a LEO constellation with an ISL network is characterized by a regular mesh topology, enabling routing procedures to be considered independently of the actual serving satellite (i.e. concealing the motion of satellites with respect to the earth from the routing procedure). Furthermore, the high level of node connectivity (typically between 2 and 6 links to the neighbouring nodes) provides several alternative paths between a given pair of satellites.
- Constant number of network nodes – routing procedures in ISL networks are based typically on the explicit knowledge of the network topology which, in the case of satellite constellation, has a constant, predefined number of network nodes in the space (satellites) and terrestrial (gateways) segments (except in the case of a node or a link failure). This property has a direct influence on the calculation of routing tables.

The above properties are incorporated in the described routing and traffic modelling techniques and procedures. Special attention is given to properties which support the development of efficient, yet not excessively complex, adaptive routing and traffic engineering techniques.

However, for the verification, validation and performance evaluation of algorithms, protocols, or whole telecommunication systems, the development of suitable traffic models, which serve as a vital input parameter in any simulation model, is of paramount importance. Thus, at the end of the chapter we are presenting the methodology for modelling global aggregate traffic comprising of four main modules. It can be used as a whole or only selected modules can be used for particular purposes connected with simulation of particular models.

Routing and traffic engineering on one side require good knowledge of the type of network and its characteristics and on the other side also of the type of traffic in the network. This is needed not only for adapting particular techniques, procedures and algorithms to the

network and traffic conditions but also for their simulation, testing and benchmarking. To this end this chapter is complemented by description of a methodology for developing a global traffic model suitable for the non-geostationary ISL networks, which consists of modules describing distribution of sources, their traffic intensity and its temporal variation, as well as traffic flow patterns.

2. Routing functions

The main task of any routing is to find suitable paths for user traffic from the source node to destination node in accordance with the traffic's service requirements and the network's service restrictions. Paths should accommodate all different types of services using different optimisation metrics (e.g. delay, bandwidth, etc.). Thus, different types of traffic can be routed over different routes. Routing functionality can be in general split in four core routing functions, (i) acquiring information about the network and user traffic state, and link cost calculation, (ii) distributing the acquired information, (iii) computing routes according to the traffic state information and chosen optimization criteria, and (iv) forwarding the user traffic along the routes to the destination node.

For each of these functions, several policies exist. Generally speaking, the selection of a given policy will impact (i) the performance of the routing protocol and (ii) the cost of running the protocol. These two aspects are dual and a careful design in the routing algorithm must achieve a suitable balance between the two. The following sub-sections will discuss the four core routing functions.

2.1 Acquiring information about the network and link cost calculation

The parameters of the link-cost metric should directly represent the fundamental network characteristics and the changing dynamics of the network status. Furthermore, they should be orthogonal to each other, in order to eliminate unnecessary redundant information and inter-dependence among the variables (Wang & Crowcroft, 1996). Depending on the composition rule we distinguish additive, multiplicative, concave and convex link-cost metrics (Wang, 1999). In additive link-cost metrics the total cost of the path is a sum of costs on every hop. Additive link costs include delay, jitter, cost and hop-count. Total cost of the path in the case of multiplicative link-cost metrics is a product of individual costs of links. A typical example of multiplicative link cost is link reliability. In concave and convex link-cost metrics the total cost of the path equals the cost on the hop with the minimum and maximum link cost respectively, and a typical example of link-cost metric is the available bandwidth.

2.1.1 Link cost for delay sensitive traffic

We show the use of the additive link-cost metric as an example for the link-cost function for the delay sensitive traffic, considering two dynamically changing parameters. The first is the intersatellite distance between neighbouring satellites, while the second is the traffic load on a particular satellite. They have a significant effect on the routing performance and are scalable with the network load and link capacity, thus being well suited for link-cost metric. (Mohorcic et al., 2004; Svigelj et al., 2004a).

The distance between satellite pairs in a non-geostationary satellite system is deterministic and can be calculated in advance. We consider this distance of a particular link l through propagation delay (T_{Pl}). Propagation delay in satellite communications is proportional to the number of hops between source and destination satellites, which could be used as a simplified cost metric or an additional criterion.

The traffic load on a particular satellite and its outgoing links is constantly changing in a random fashion, thus it needs to be estimated in real-time. To estimate the traffic load on particular link we can use two parameters. It can be estimated through the queuing delay, which reflects the past values of traffic load, or expected queuing delay, which estimates the future value of queuing delay in a given outgoing queue. In addition, both parameters can be improved with additional functions (i.e. exponential forgetting function, exponential smoothing function), which are described in the following subsections. Thus, in general the link costs (LC_l) for delay sensitive traffic on the link l at time t_i are calculated using Equation (1) at the end of each routing table update interval. It includes the propagation delay (T_{Pl}) and traffic load represented by (T_{Ql}).

$$LC_l(t_i) = T_{Pl}(t_i) + T_{Ql}(t_i) \tag{1}$$

2.1.1.1 Link cost based on the queuing delay enhanced with Exponential forgetting function EFF

In this case we monitor the traffic load on a satellite through the packet queuing delay (T_{ql}) at the respective port of the node, which is directly proportional to the traffic load on the selected outgoing link l as shown in Equation (2), where L_r denotes the length of the r^{th} packet in outgoing queue and C_l is the capacity of the link l

$$T_{ql} = \frac{\sum_r L_r}{C_l} \tag{2}$$

Due to variation of these queuing delays, the queuing delay value T_{Ql}, considered in the link-cost function, is periodically estimated using a fixed-size window exponential forgetting function EFF(n, χ, T_{ql}) on a set of the last n values of packet queuing delay collected in a given time interval (i.e. $T_{ql}[n]$ being the last collected value, and the other values considered being $T_{ql}[n\text{-}1]$,..., $T_{ql}[1]$). In the EFF function, n (the depth of the function) denotes the number of memory cells in the circular register. If the number of collected T_{ql} values m is smaller than n, then only these values are considered in the EFF function. Furthermore, as shown in Equation (3), a forgetting factor, $\chi \in (0, 1)$, is introduced to make the more recent T_{ql} values more significant in calculating T_{Ql}.

$$T_{Ql} = EFF(n, \chi, T_{ql}) = \begin{cases} (1-\chi)\cdot \sum_{r=0}^{m-1} \chi^r \cdot T_{ql}[m-r] & \text{for } m < n \\ (1-\chi)\cdot \sum_{r=0}^{m-1} \chi^r \cdot T_{ql}[n-r] & \text{for } m \geq n \end{cases} \tag{3}$$

2.1.1.2 Link cost based on expected queuing delay enhanced with Exponential Smoothing Link-Cost Function

In the case of using expected queuing delay in the assessment of the traffic load, we monitor the outgoing queues of particular traffic. A packet entering a given output queue at time t will have the expected queuing delay, T_{exp}, given by Equation (4), where L_{av} is the average packet length, C the link capacity, and $n(t)$ the number of packets in the queue.

$$T_{exp}(t) = n(t) \cdot \frac{L_{av}}{C} \tag{4}$$

Calculation of the expected queuing delay does not require any distribution of link status between neighbouring nodes, and has the advantage of fast response to congestions on the link. However, for calculation of pre-computed routing tables the average expected queuing delay T_{exp_av} has to be determined using Equation (5) at the end of each update interval T_I starting at time t_S. This average expected queuing delay could subsequently be already used as a link-cost metric parameter T_{Ql}, as shown in Equation (6), which expresses traffic load on the link.

$$T_{exp\_av}(t_S + T_I) = \frac{1}{T_I} \cdot \int_{t_S}^{t_S+T_I} n(t) \cdot \frac{L_{av}}{C} \cdot dt \tag{5}$$

$$T_{Ql}(t_i) = T_{exp\_av}(t_S + T_I) \tag{6}$$

The consideration of link load in the link cost calculation, and consequently in route computation, may cause traffic load oscillations between alternative paths in the network (Bertsekas & Gallager, 1987). In particular, routing of packets along a given path increases the cost of used links. At the end of routing update interval this information is fed back to the routing algorithm, which chooses for the next routing update interval an alternative path. In extreme cases this may result in complete redirection of traffic load to alternative paths, eventually leading to traffic load oscillation between the two alternative paths in consecutive routing tables and hence routing instability. In ISL networks for instance traffic load oscillations impose a particular effect on delay sensitive traffic, as there are many alternative paths between a given pair of satellites with similar delays. Oscillations are especially inconvenient under heavy traffic load conditions, where the impact of traffic load parameter on the link cost is much higher than that of the propagation delay T_P. Under such conditions oscillations lead to congestion on particular links, which significantly degrades routing performance. In addition, the oscillations of traffic load have also a great impact on triggered signalling, where the signalling load depends on a significant change of link cost. In order to introduce the triggered signalling, the reduction of the oscillation of traffic load and consequently the oscillation of link cost is inevitable. Smoothing of the link cost on a particular link can be done in two ways:

- Directly by modifying the link cost on particular link with a suitable smoothing function.

- Indirectly by using advanced forwarding policies, which send traffic also along the alternative paths and distribute traffic more evenly on the first and the second shortest paths and consequently smooth-out the link cost. (see section 2.4.)

To reduce the oscillations one can use an exponential smoothing link-cost function, which iteratively calculates the traffic load parameter T_{Ql} from its previous values according to Equation (7). The influence of the previous value is regulated with a parameter k, defined between 0 and 1 ($k \in [0,1]$), while the initial value for the parameter T_{Ql} is set to 0.

$$\begin{aligned} T_{Ql}(t_0) &= 0 \\ T_{Ql}(t_i) &= \left[T_{exp\_av}(t_i) - T_{Ql}(t_{i-1})\right] \cdot k + T_{Ql}(t_{i-1}) = \\ &= k \cdot T_{exp\_av}(t_i) + (1-k) \cdot T_{Ql}(t_{i-1}) \end{aligned} \tag{7}$$

Taking into account this parameter, the cost of a given link l is calculated using Equation (1). If k equals 1, there is no influence of previous values on current link cost and Equation (7) transforms to Equation (6). On the other hand, if k equals 0, only propagation delay is considered in link cost calculation, which leads to traffic insensitive routing.

One of the drawbacks of the exponential smoothing link-cost function is that it takes into account in each iteration all previous values of parameter T_{exp_av}. The value of T_Q as a function of n previous values of T_{exp_av} is given in Equation (8). It can be seen that the impact of previous values of T_{exp_av} decreases exponentially with increasing value of n.

$$\begin{aligned} T_{Ql}(t_n) = k \cdot ((1-k)^0 \cdot T_{exp\_av}(t_n) + (1-k)^1 \cdot T_{exp\_av}(t_{n-1}) + \\ + (1-k)^2 \cdot T_{exp\_av}(t_{n-2}) + \ldots + (1-k)^{n-1} \cdot T_{exp\_av}(t_1)) \end{aligned} \tag{8}$$

The main goal of the exponential smoothing link-cost function, which tends to suppress the traffic load oscillations, is that the link cost should reflect the actual traversing traffic flow and the traffic intensity of the region served by the satellite, and not the instantaneous fluctuations of traffic load due to oscillation. In such manner exponential smoothing algorithm promises more evenly distribution of traffic load between links and consequently a better performance for different traffic types. Furthermore it ensures, that in a lightly loaded network, the routing performance is not decreased, while it is notably enhanced in heavily loaded network. A more exhaustive explanation of exponential smoothing link cost function and optimum definition of parameter k is given in (Svigelj et al., 2004a).

2.1.1.3 Weighted delay calculation

The relative impacts of traffic load and propagation delay on the link cost are linearly regulated with a traffic weight factor (TWF_l) and a propagation delay weight factor ($PDWF_l$), respectively, as shown in Equation (9) defining weighted delay (WD_l) on the link l. This allows biasing of link cost towards shortest-path routes ($PDWF_l > TWF_l$) or towards least loaded but slightly longer routes ($PDWF_l < TWF_l$).

$$WD_l = PDWF_l \cdot T_{Pl} + TWF_l \cdot T_{Ql} \tag{9}$$

In general, as indicated in Equation (9), different weights can be used on different links. In a non-geostationary satellite system, however, satellites are continuously revolving around the rotating earth, so weights cannot be optimized for the traffic load of certain regions but should either be fixed or should adapt to the conditions in a given region. The later gives opportunity for further optimisation using some traffic aware heuristic approach.

Weighted delay on the link, as given by Equation (9), can already be used as a simple continuous link-cost function with a linear relation between both metrics. In general, however, a more sophisticated link-cost function should be able to control the relative cost of heavily loaded links with respect to lightly loaded links. This can be accomplished by a non-linear link-cost function, such as an exponentially growing function with exponent α, as given in Equation (10), where WD_L and WD_U represent lower and upper boundary values of weighted delay on the links respectively.

$$LC_l = \left(\frac{WD_l - WD_L}{WD_U - WD_L} \right)^{\alpha} + \frac{WD_L}{WD_U} \tag{10}$$

The first term in Equation (10) represents the normalised dynamically changing link cost according to variation of propagation delay (e.g. ISL length) and traffic load (e.g. queuing delay). Since it is not suitable that link cost be zero, which can cause high oscillations, a small constant (WD_L/WD_U) is added to the normalised term of the link-cost function. This constant represents the normalised cost of the shortest link without any traffic load. When $a = 0$ a link-cost function has no influence on the routing algorithm, and path selection reduces to cost-independent routing (i.e. minimum hop count routing), while with $a = 1$ it selects a path with the minimum sum of link costs. Exponent values larger than 1 ($a > 1$) tend to eliminate heavily loaded (high cost) links from consideration, while exponent values smaller than 1 ($a < 1$) tend to preserve lightly loaded links. Combining Equations (9) and(10), the link cost for the delay sensitive traffic, which takes into consideration delay on the link, is calculated as given by Equation (11).

$$LC_l = \left(\frac{PDWF_l \cdot T_{Pl} + TWF_l \cdot T_{Ql} - WD_L}{WD_U - WD_L} \right)^{\alpha} + \frac{WD_L}{WD_U} \tag{11}$$

2.1.1.4 Discretization

Regardless of the selected link-cost function the calculated link cost needs to be distributed throughout the network and stored in nodes for the subsequent calculation of new routing tables. In order to reduce computation effort and memory requirements, routing algorithms have been proposed that perform path selection on a small set of discrete link-cost levels. In these algorithms the appropriate number of link-cost levels needs to be defined to balance between the accuracy and computational complexity.

Equation (13) represents a suitable function, which converts the continuous link-cost function, given in Equation (12), to L discrete levels denoted as C_{Dl} in the range between 0 and 1. In this link-cost function the minimum and maximum value for weighted delay are used, WD_{min} and WD_{max}. Any link with weighted delay below WD_{min} is assigned the minimum cost $1/L$, while links with weighted delay higher than WD_{max} have link cost set to 1.

$$C_i = WD_i^{\alpha} \tag{12}$$

$$C_{Dl} = \begin{cases} \dfrac{1}{L} & WD_l < WD_{min} \\ \dfrac{\left\lceil \left(\dfrac{WD_l - WD_{min}}{WD_{max} - WD_{min}} \right)^a \cdot (L-1) \right\rceil + 1}{L} & WD_{min} < WD_l < WD_{max} \\ 1 & WD_l \geq WD_{max} \end{cases} \tag{13}$$

2.1.2 Link cost function for the throughput sensitive traffic

The most suitable optimization parameter for the throughput sensitive traffic, on the other hand, is the available bandwidth on the link. Thus, on each link the lengths of the traversing packets are monitored between consecutive routing table updates, and the link utilization (LU_l) is calculated according to Equation (14), where L_r denotes the length of the r^{th} traversing packet. The selected time interval between consecutive calculations of the sum of the packet lengths was equal to the routing table update interval T_I starting at time t_S.

$$LU_l(t_S + T_I) = \frac{\sum_r L_r}{T_I \cdot C_l} \tag{14}$$

The link-cost metric for the throughput sensitive traffic is a typical concave metric. The optimization problem is to find the paths with the maximum available bandwidth and, as an additional constraint, with minimum hop count, which minimizes the use of resources in the network. Thus, the link cost for throughput sensitive traffic is the normalized available bandwidth on the link, calculated at the end of the routing table update interval according to Equation (15).

$$LC_l(t_i) = 1 - LU_l(t_S + T_I) \tag{15}$$

2.2 Distributing the acquired information – signalling

Before the routes are calculated the information about network state should be distributed between nodes. An effective signalling scheme must achieve a trade-off between (a) bandwidth consumed for signalling information (b) computing and memory capacity dedicated to signalling processing and (c) improvement of the routing decisions due to the presence of signalling information (Franck & Maral, 2002a). Signalling is subdivided in two families: unsolicited and on-demand signalling. The following subsections detail these two families.

2.2.1 Unsolicited signalling

Unsolicited signalling is similar to unsolicited mail ads. Nodes receive at given time intervals information about the state of the other nodes. Conversely, nodes broadcast in the

network information about their own state. Because a node has no control of the time it receives state information, the information might be non-topical once used for route computation. Non topical information is undesirable since it introduces a discrepancy between what is known and what the reality is. This is of particular importance for those systems which incorporate non-permanent links. Non topical information results in inaccurate and possibly poor routing decisions. Unsolicited signalling is further subdivided into periodic and triggered signalling.

Periodic signalling works by having each node broadcasting state information every p units of time, p being the broadcast period. It is not required for the broadcast period be equal for all nodes, however, it is practical to do so because (a) all nodes run the same software (b) it avoids discrepancies in the topicality of state information. Since the quality of routing decisions depends on how topical the state information is, it is expected that increasing the broadcast period results in increasing the connection blocking probability. On the other hand, increasing the broadcast period helps to keep the signalling traffic low. Periodic signalling supports easy dimensioning since the amount of signalling traffic does not depend on the amount of traffic flowing in the network and therefore can be quantified analytically. Unfortunately, this interesting characteristic is also a drawback: if the state of a node does not change during the whole broadcast period, the next broadcast will take place, regardless of whether it is useful or not. Likewise, some important state change might occur in the middle of the broadcast period without any chance for these changes to be advertised prior to the next broadcast. For these reasons, triggered signalling is worth investigating.

Instead of broadcasting periodically, the node using **triggered signalling** permanently monitors its state and initiates a broadcast upon a significant change of its state (threshold function). This approach is supposed to alleviate signalling traffic, holding down useless broadcasts. Triggered updates for instance are used for Routing Information Protocol (RIP). Unfortunately, triggered signalling has two down sides. First, while periodic signalling does not depend on the actual content of state information, triggered signalling must be aware of the semantics of the state information to define what a significant state change is. Second, the amount of signalling traffic generated depends on the characteristics of the traffic load in the constellation. It does not depend on the amount of data traffic but rather on the traffic variations in the nodes and links. Since routing impacts how traffic is distributed in the network, the behaviours of routing and triggered signalling are tightly interlaced. Triggered signalling can be further sub-divided in additional versions depending on the chosen threshold function.

In networks there are two changing parameters, which have the impact on the link cost: propagation delay between neighbouring nodes and traffic load. The first can be computed in advance in each node, so it can be eliminated from signalling information. For delay sensitive traffic the new value of T_{Ql} is broadcasted only if the value exceeds predefined threshold (Svigelj et al., 2012). If T_{Ql} does not exceed the threshold, only the propagation delay is used as a link cost in routes calculation. In the case of throughput sensitive traffic the link cost is broadcasted only if LC_l is lower than threshold (i.e. the available bandwidth is lower than threshold), otherwise value 1 (i.e. empty link) is used in routes calculation.

With an appropriate selection of thresholds the signalling load can be significantly reduced, especially for nodes, which has no intensive traffic. To omit the impact of oscillations of the

link costs the triggered signalling can be used in a combination with exponential smoothing link-cost function or adaptive forwarding.

2.2.2 On-demand signalling

Compared to unsolicited signalling, on-demand signalling works the other way around. When a node (called the requesting node) requires state information, it queries the other nodes (called the serving nodes) for this information. Thus, on-demand signalling yields the state information as recent as possible, with expected benefit for the routing decisions. Furthermore, the type of state information which is queried (e.g. capacity or buffer occupancy) may vary according to the type of route that must be computed. On the other hand, since the signalling procedure is triggered for each route computation, the amount of traffic generated by on-demand signalling is likely to be higher than with unsolicited signalling. Additionally, the requesting node has to gather complete information before initiating the route computation. On-demand signalling is more convenient for connection oriented networks, where the source node requests the network state information from other nodes before setting up a connection and then the route to destination node is computed. As the number of packets during a signalling session is high, additional mechanisms (caching, snooping) have to be devised, in order to limit the number of signalling packets (Franck & Maral, 2002a).

2.3 Computing routes

In the case of per-hop packet-switched routing routes cannot be computed on demand. Instead, routing tables are pre-computed for all nodes periodically or in response to a significant change in link costs, thus defining routing update intervals. Link-cost metrics for the delay sensitive traffic are typical additive metrics, and thus the shortest routes are typically calculated using the Dijkstra algorithm. The main feature of an additive metric is that the total cost for any path is a sum of costs of individual links.

On the other hand, the link cost for the throughput sensitive traffic is a concave metric. Thus, the total cost for any path equals the one on the link with minimum cost. A typical optimization criterion for the throughput sensitive traffic is to find the paths within minimum hop count with the maximum available bandwidth. Minimum hop count is an additional constraint, which is used to minimize the use of resources. The Bellman-Ford shortest path algorithm is well suited to compute paths of the maximum available bandwidths within a minimum hop count. It is a property of the Bellman-Ford algorithm that, at its h^{th} iteration, it identifies the optimal path (in our context the path with the maximum available bandwidth) between the source and each destination not more than h hops away. In other words, because the Bellman-Ford algorithm progresses by increasing the hop count, it provides the hop count of a path as a side result, which can be used as a second optimization criterion.

Regardless of the type of traffic the second shortest path with disjoint first link can be calculated by eliminating the first link on the shortest route (i.e. LC_l is set to infinity for delay sensitive traffic and to 0 in the case of throughput sensitive traffic) and using Dijkstra and Bellman Ford algorithm on such modified network. The alternative paths are used in the case of adaptive forwarding.

2.4 Forwarding the user traffic

In the route execution phase packets are forwarded on outgoing links to the next node along the path according to most recently calculated routing tables. In particular, packets are placed into an appropriate first in first out (FIFO) queue with a suitable scheduler according to the traffic type they belong to and according to the selected forwarding policy.

2.4.1 Static forwarding

Two representatives of static forwarding policies originally developed for regular network topologies, such as exhibited by ISL networks, are alternate link routing with deflection in the source node (ALR-S) and alternate link routing with deflection in all nodes (ALR-A) (Mohorcic et al., 2000, 2001). Both policies are based on an iterative calculation of routing algorithm for determining alternative routes between satellite pairs. An additional restriction considered in static forwarding policies is that the alternative routes must consist of the same (i.e., minimum) number of hops, with a different link for the first hop. Such alternative routes with the same number of hops guarantee that the propagation delay increase for the second-choice route is kept within a well-defined limit.

After determination of alternative routes with the same number of hops between each pair of nodes (satellites) the selected forwarding policy decides which packets are forwarded along each of these routes. Different forwarding policies are depicted in Fig. 1

According to the routing table given in Table 1, the SPR policy is only forwarding user traffic along the shortest routes. This leads to very non-uniform traffic load particularly on links (A-D, B-E, and C-F).

	Next hops on the route to satellite F and the cost of the route					
From	Shortest route		Second shortest route		Third shortest route	
Satellite A	D, E, F	14	B, E, F	15	B, C, F	16
Satellite B	E, F	10	C, F	11	/	/
Satellite C	F	6	/	/	/	/
Satellite D	E, F	10	/	/	/	/
Satellite E	F	5	/	/	/	/

Table 1. Alternative paths to Satellite F with the same minimum number of hops.

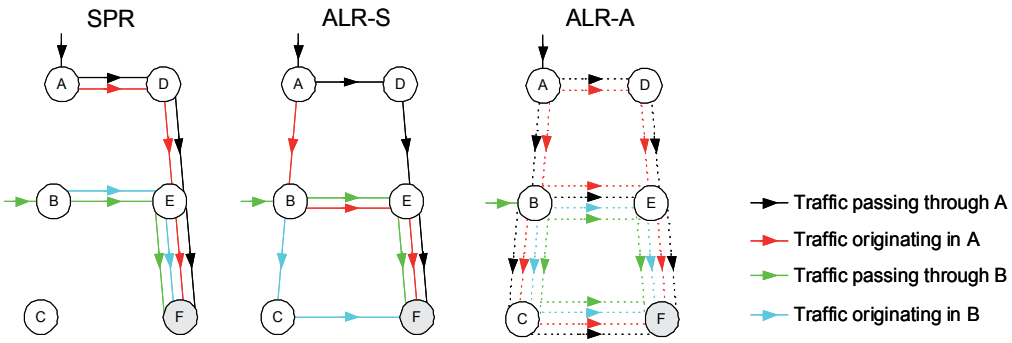


Fig. 1. Path selection with different forwarding policies.

The ALR-S policy ensures a more uniform distribution of traffic load over the network, as it distinguishes between the packets passing through a particular node and the packets that are originating in that node. Packets originating in a particular node are forwarded on the link of the second shortest route (e.g. from A to F via B, from B to F via C), while packets passing through the node are forwarded on the link of the shortest route (e.g. through A to F via D, through B to F via E). By using the second-choice route only for originating packets, the delay is increased with respect to the shortest route only on the first hop, hence the increase in delay does not accumulate for the packets with a large number of hops. Between the consecutive updates of routing tables, all packets between a given pair of nodes follow the same route. Thus, ALR-S policy maintains the correct sequence of the packets within the routing interval, the same as the SPR forwarding policy.

The ALR-A policy promises an even more uniform distribution of traffic load and thus further improvement of link utilisation by alternating between the shortest and the second shortest route regardless of the packet origination node (this is denoted in Fig. 1 by dashed lines). However, packets belonging to the same session can be forwarded along different routes even within one routing table update interval, thus additional buffering is required in the destination nodes to re-order terminated packets and obtain the correct sequence.

The static forwarding policies, such as ALR-S and ALR-A, distribute packets according to a pre-selected rule. They allow significant reduction of traffic load fluctuation between links, however they do not adapt to the actual traffic load on alternative routes.

2.4.2 Adaptive forwarding

In contrast to static forwarding an adaptive forwarding policy has to take into account the link status information to support the selection of the most appropriate between the alternative outgoing links on the route to the destination. An example of such approach is adaptive forwarding policy based on local information about the link load (Svigelj et al, 2003, 2004b; Mohorcic et al. 2004). This policy selects the most suitable outgoing link taking into account routing tables with alternative routes, calculated using link costs obtained during the previous routing update interval, and current local information on the link status.

In particular, for delay sensitive traffic local information can be based on the expected queuing delay as defined in Equation (7). The expected queuing delay for a particular link can be calculated locally and does not require any information distribution between neighbouring nodes, thus enabling a very fast response to congestion on the link. Depending on this local information, packets are forwarded on the shortest or on the alternative second shortest path. The alternative second shortest path is used only if it has the same or a smaller number of hops (h) to the destination and if the expected queuing delay in the outgoing queue on the shortest path (T_{exp1}) is more than a given threshold Δ_{tr}^{D} (where D is denoting delay sensitive traffic) higher than the expected queuing delay in the outgoing queue on the second shortest path (T_{exp2}). This condition for selecting the alternative second shortest path is given in Equation (16). Different threshold values can be used for different traffic types.

$$(h_2 \le h_1) \wedge (T_{\exp 1}(t) - T_{\exp 2}(t) < \Delta_{tr}^{D} \tag{16}$$

For the throughput sensitive traffic we monitor the number of packets in outgoing queues (n). The alternative second shortest path is used only if it has the same or a smaller number of hops (h) to the destination and if the number of packets (n) in the outgoing queue on the shortest path (n_1) is more than a given threshold Δ_{tr}^{T} (where T is denoting throughput sensitive traffic) higher than the number of packets in the outgoing queue on the alternative path (n_2), as given in Equation (17).

$$(h_2 \le h_1) \wedge (n_1(t) - n_2(t) < \Delta_{tr}^{T} \tag{17}$$

The significance of the threshold is that it regulates distribution of traffic between alternative paths based on local information about the link status, and thus differentiates between lightly and heavily loaded nodes. The higher the threshold value the more congested the shortest path needs to be to allow forwarding along the alternative second shortest path. In the extreme, setting the threshold value to infinity prevents forwarding along the second shortest path (i.e. adaptive forwarding deteriorates to SPR), while no threshold (i.e. $\Delta_{tr}^{T} = 0$) means that packets are forwarded along the second shortest path as soon as the expected queuing delay for the corresponding link is smaller than the one on the shortest path.

Routing with the proposed adaptive forwarding promises more uniform distribution of traffic load between links and the possibility to react quickly to link failure. However, packets belonging to the same session can be forwarded along different routes, even within the same routing update interval, so additional buffering is required in destination nodes to reorder terminated packets and obtain the correct sequence.

3. Traffic modelling for global networks

As we have shown in previous section, the general routing and traffic engineering functions consist of many different algorithms, methods and policies that need to be carefully selected and adapted to the particular network characteristics as well as types of traffic to be used in the network. Clearly, the more dynamic and non-regular the network and the more different types of traffic, the more demanding is the task of optimising network performance, requiring good understanding of the fundamental network operating conditions and the traffic characteristics. The later largely affect the performance of routing and traffic engineering, typically requiring appropriate traffic models to be used in simulating, testing and benchmarking different routing and traffic engineering solutions. In the following a methodology is described for developing a global traffic model suitable for supporting the dimensioning and computer simulations of various procedures in the global networks but focusing in particular on the non-geostationary ISL networks, which are well suited for supporting asymmetric applications such as data, audio and video streaming, bulk data transfer, and multimedia applications with limited interactivity, as well as the broadband access to Internet services beyond densely populated areas. Such traffic models are an important input to network dimensioning tasks (Werner et al., 2001) as well as to simulators devoted to the performance evaluation of particular network functions such as routing and traffic engineering (Mohorcic et al., 2001, 20021, Svigelj et al., 2004a).

A typical multimedia application contains a mix of packets from various sources. Purely mathematical traffic generators cannot capture the traffic characteristics of such applications in real networks to the extent that would allow detailed performance evaluation of the

network. Hence, the applicability of traffic analysis based on mathematical tractability is diminishing, while the importance of computer simulation has grown considerably, but poses different requirements for traffic source models (Ryu, 1999). A suitable traffic source model should represent real traffic, while the possibility of mathematical description is less important. In global non-geostationary satellite network traffic source model needs to be complemented by a suitable model of other elementary phenomena causing traffic dynamics, i.e. geographical distribution of traffic sources and destinations, temporal variation of traffic load and traffic flow patterns between different geographical regions.

In the following the approach to modelling global aggregate traffic intensity is described, in particular useful for the dimensioning of satellite networks and computer simulations of various procedures in the ISL network segment, including routing and traffic engineering.

The model is highly parameterized and consists of four main modules:

- module for global distribution of traffic sources and destinations;
- module for temporal variations of traffic sources' intensity;
- module describing the traffic flow patterns between regions; and
- module describing statistical behaviour of aggregated traffic sources.

3.1 Module for global distribution of traffic sources and destinations

The module for global distribution of traffic sources and destinations should support the representation of an arbitrary distribution.

A simple representative of a geographically dependent source/ destination distribution assumes homogeneous distribution over the landmasses, considering continents and major islands (called landmass distribution), while traffic intensity above the oceans equals 0 (Mohorcic et al., 2002b). More realistic source/destination distributions should reflect the geographic distribution of traffic intensity, which is related to several techno-economic factors including the population density and distribution, the existing telecommunication infrastructure, industrial development, service penetration and acceptance level, gross domestic product (GDP) in a given region, and pricing of services and terminals (Werner & Maral, 1997, Hu & Sheriff, 1997, Werner & Lutz 1998). Thus, the estimation of traffic distribution in the yet non-existing system demands a good understanding of the types of services and applications that will be supported by the network. Furthermore, it should also consider attractiveness of particular services for potential users, which in turn depends also on different socio-economic factors.

The methodology for estimating the market distribution for different terminal classes, i.e. lap-top, briefcase and hand-held, is reported in (Hu & Sheriff, 1998) Essentially, countries over the globe are categorized into three different bands according to their annual GDP per capita: low (less than 6 kEuro), medium (between 6 kEuro and 22 kEuro) and high (greater than 22 kEuro). A yearly growth for GDP per capita for each country is then predicted by linearly extrapolating historical data. This, together with the tariff of a particular service and a predicted market saturation value, is used to determine the yearly service take-up for each country via the logistic model. The yearly service penetration for each country is estimated by multiplying the predicted yearly gross potential market with the yearly take-up (Mohorcic et al., 2003).

Taking into account techno-economic and socio-economic factors and the above methodology, we can define different non-homogeneous geographic-dependent distributions taking into account a more realistic distribution of sources and destinations for provisioning of the particular types of service. Such geographic-dependent distributions are typically based on statistical data provided on the level of countries, and only for some larger countries also on the level of states and territories. In addition to limitations of data availability, we also face the problem of the accuracy of its representation, which depends on the granularity of the model and on the assumption regarding the source/destination distribution within the smallest geographical unit (i.e. country). The simplest approach in country-based non-homogeneous geographic-dependent distributions assumes that a nation's subscribers are evenly distributed over the country. The weakness of this approach is representation of traffic demand in large countries spanning several units of geographical granularity. In determining the distribution, different levels of geographical granularity may be adopted; however, in order to be able to individually represent also small countries, the geographical granularity should be in the range of those small countries. In (Mohorcic et al., 2003), a traffic grid of dimension 180° × 360° has been generated in steps of 1° in both latitude and longitude directions.

3.2 Module for temporal variations of traffic sources' intensity

Temporal variation of traffic load in a non-geostationary satellite system is caused by daily variation of traffic load due to the local time of day and geographical variation of this daily load behaviour according to geographical time zones. Both are considered in the module for temporal variation of traffic load, which actually mimics the geographically dependent daily behaviour of users. Daily variation can be taken into account with an appropriate daily user profile curve (for average or for local users). An example of such a daily user profile curve is shown in Fig. 2. For geographical time zones a simplified model can be considered, which increments the local hour every 15 degrees longitude eastward from the GMT.

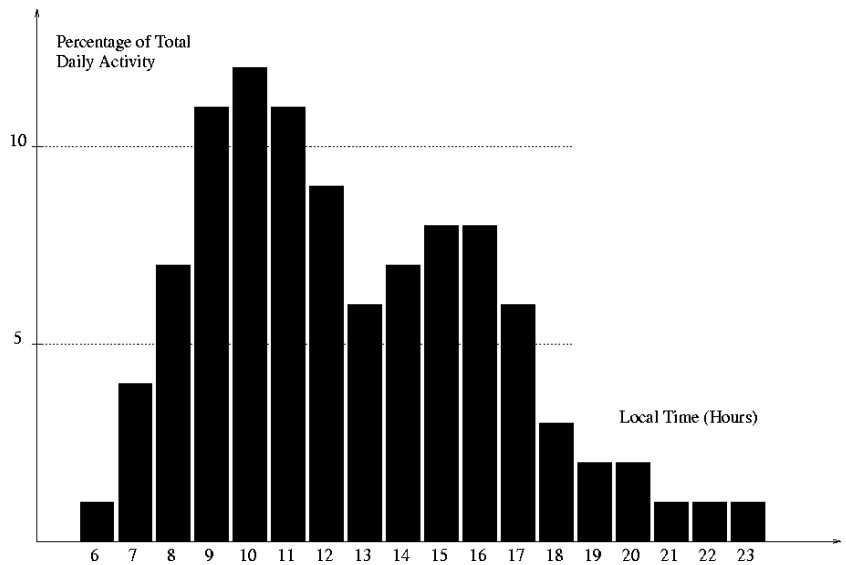


Fig. 2. Daily user profile curve.

An alternative approach defines temporal variation of traffic load in conjunction with the global distribution of traffic sources and destinations, which inherently takes into account geographical time zones. An example of relative traffic intensity considering distribution of traffic sources and destinations combined with temporal variation of traffic load is depicted in Fig. 3, where traffic intensity is normalised to the highest value (i.e. the maximum value of normalized traffic load equals 1, but for better visualization we bounded the z-axis in Fig. 3 to 0.3). The traffic intensity is generated by assuming that a single session is established per day per user and that each session on average lasts for about 2 minutes.

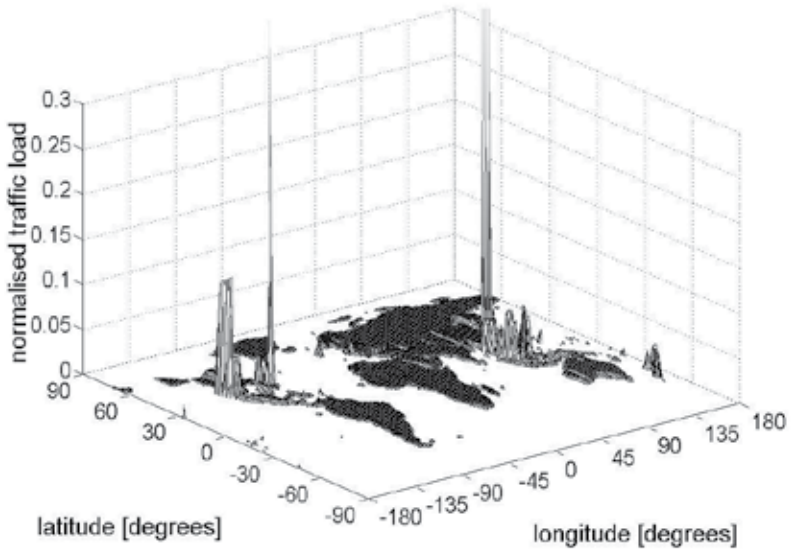


Fig. 3. Global distribution and activity of traffic sources and destinations at midnight GMT.

Another contribution to temporal variation of traffic load in non-geostationary ISL networks in addition to user activity dynamics is the rapidly changing satellite visibility, and consequently active users' coverage, on the ground. To a certain extent this temporal variation as well as multiple visibility of satellites can be captured with a serving satellite selection scheme. Implementing a satellite selection scheme in case of multiple visibility has two aspects. For fixed earth stations line-of-sight conditions are assumed, so that the serving satellite can be determined according to a simple deterministic rule, e.g., maximum elevation satellite. For mobile earth stations, the stochastic feature of unexpected handover situations due to propagation impairments can be considered through the shares of traffic on alternative satellites also estimated according to a simple rule (e.g., equal sharing between all satellites above the minimum elevation) or using a simple formula (e.g., shares are a function of the elevation angle of each alternative satellite as one main indicator for channel availability).

3.3 Module describing the traffic flow patterns between regions

This module assigns traffic flow destinations using a traffic flow pattern resembling the flow characteristic between different regions. Interregional patterns should be defined at least on the level of the Earth's six continental regions shown in Fig. 4, similarly as in (Werner & Maral,

1997), but preferably on a smaller scale between countries/territories. In a destination region, the traffic can be divided among the satellites proportionally to their coverage of that region.

Customized traffic flow patterns should be based on the density distribution of sources and/or destinations for the selected type of service.

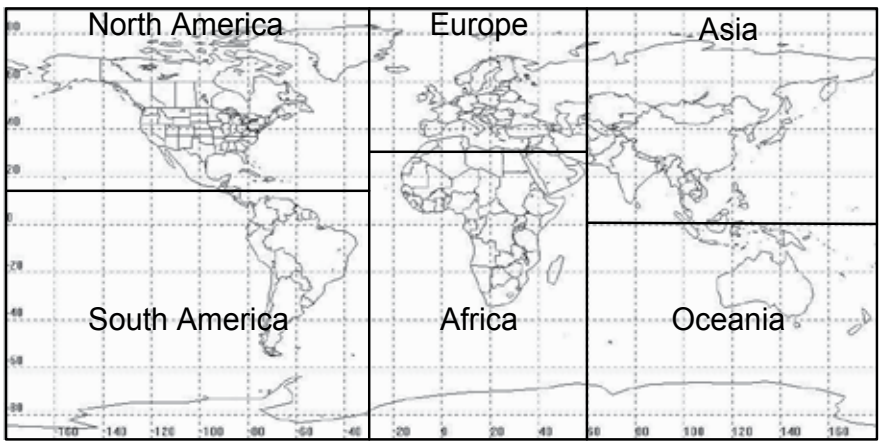


Fig. 4. Geographical division of six source/destination regions.

3.4 Module describing statistical behaviour of aggregated traffic sources

The fourth module concerns modelling of the aggregated traffic sources. In particular, the module comprises of suitable aggregate traffic source generator, which is modulated by the normalized cumulative traffic on each satellite obtained from distribution of traffic sources and destinations and temporal variation of traffic sources' intensity. Thus data packets are actually generated considering the relative traffic intensity experienced by a particular satellite in its coverage area, while taking into account the statistics of the selected aggregate traffic source model.

Ideally, the traffic source model should capture the essential characteristics of traffic that have significant impact on network performance with only a small number of parameters, and should allow fast generation of packets. Among the most important traffic characteristics for circuit switched networks are the connection duration distribution and the average number of connection requests per time unit. By contrast, in the case of packet switched networks, traffic characteristics are given typically by packet lengths and packet inter-arrival times (in the form of distributions or histograms), burstiness, moments, autocorrelations, and scaling (including long-range dependence, self-similarity, and multifractals). For generating cumulative traffic load on a particular satellite, the traffic source generator should model an aggregate traffic of many sources overlaid with the effect of a multiple access scheme, which is expected to significantly shape source traffic originating from single or multiplexed ground terminal applications due to the uplink resource management and traffic scheduling.

One approach for modelling aggregate traffic sources is by using traces of real traffic. Trace-driven traffic generators are recommended for model validation, but suffer from two

drawbacks: firstly, the traffic generator can only reproduce something that has happened in the past, and secondly, there is seldom enough data to generate all possible scenarios, since the extreme situations are particularly hard to capture. In the case of satellite networks with no appropriate system to obtain the traffic traces, the use of traces is even more inconvenient.

An alternative approach, increasingly popular in the field of research, is to base the modelling of traffic sources on empirical distributions obtained by measurement from real traffic traces. The measurements can be performed on different segments of real networks, i.e. in the backbone network or in the access segment. In order to generate cumulative traffic load representing an aggregate of many individual traffic sources in the coverage area of the satellite, the traffic properties have to be extracted from a representative aggregate traffic trace (Svigelj at al., 2004a), such as a real traffic trace captured on the 622 Mbit/s backbone Internet link carrying 80 Mbit/s traffic (Micheel, 2002). The selected traffic trace comprises aggregate traffic from a large number of individual sources. Such traffic trace resembles the traffic load experienced by a satellite, both from numerous traffic sources within its coverage area, and from aggregate flows transferred over broadband intersatellite links. A suitable traffic source model, which resembles IP traffic in the backbone network, can already be built by reproducing some of the first order statistical properties of the real traffic trace that have major impact on network performance, e.g. inter-arrival time and packet length distribution. A simple traffic generator can be developed using a look-up table with normalized values, which allows packet inter-arrival time and packet length values to be scaled, so as to achieve the desired total traffic load. Distributions of packet inter-arrival time and packet length obtained with such a traffic generator are depicted in Fig. 5 and Fig. 6 respectively. The main advantage of traffic sources, whose distributions conform to those obtained by measurements of real traffic, is that they are relatively simple to implement and allow high flexibility.

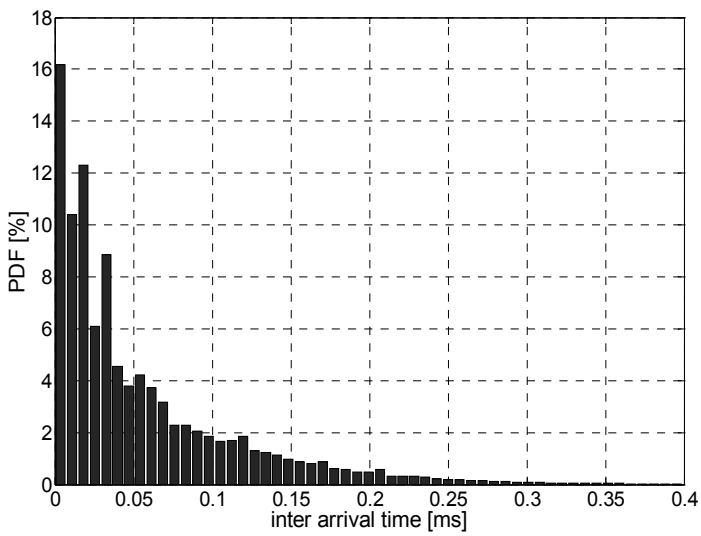


Fig. 5. Packet inter-arrival time distribution obtained with empirical traffic generator.

For the more accurate prediction of the behaviour of the traffic source exhibiting long-range dependence, the traffic model requires detailed modelling of also the second order statistics of the packet arrival process. The accurate fitting of modelled traffic to the traffic trace can be achieved using modelling process with a discrete-time batch Markovian arrival process that jointly characterizes the packet arrival process and the packet length distribution (Salvador et al., 2004). Such modelling allows very close fitting of the auto-covariance, the marginal distribution and the queuing behaviour of measured traces.

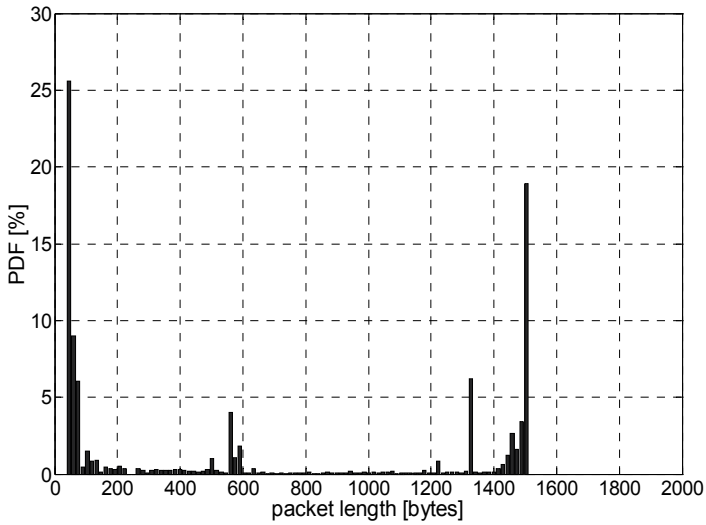


Fig. 6. Packet length distribution obtained with empirical traffic generator.

The potential drawback of traffic sources based on real traffic traces is that the empirically obtained traffic properties (i.e. obtained from the aggregated traffic on the backbone Internet link in this particular example) may not be suitably representative for the system under consideration, so it can sometimes deviate considerably from real situations and lead to incorrect conclusions.

In addition to traffic sources based on traffic traces (directly or via statistical distributions) traffic sources can also be implemented in classical way with pure mathematical distributions such as Poisson, Uniform, Self-Similar, etc. Although such mathematically tractable traffic sources never fully resemble the characteristics of real traffic, they can serve as a reference point to compare simulation results obtained with different scenarios, however they should exhibit the same values of first order statistic (i.e. mean inter-arrival time and average packet length) as obtained from traces.

In the case of supporting different levels of services, packets belonging to different types of traffic (e.g. real time, high throughput, best effort) should be generated using different traffic source models, which should reproduce statistical properties of that particular traffic. However, as different services and applications will generate different traffic intensity depending on regions and users' habits, also separate traffic flow patterns will have to be developed for different types of traffic, to be used in conjunction with different traffic source generators.

3.5 Global aggregate traffic intensity model

Integration of individual modules in the global aggregate traffic intensity model is schematically illustrated in Fig. 7. Instead of simulating individual sources and destinations, a geographic distribution of relative traffic source intensity is calculated for any location on the surface of the Earth. The cumulative traffic intensity of sources within its coverage area are mapped to the currently serving satellite. Satellite footprint coverage areas on the Earth, overlaid over geographic distribution of traffic sources and destinations, are identified from the satellite positions in a given moment.

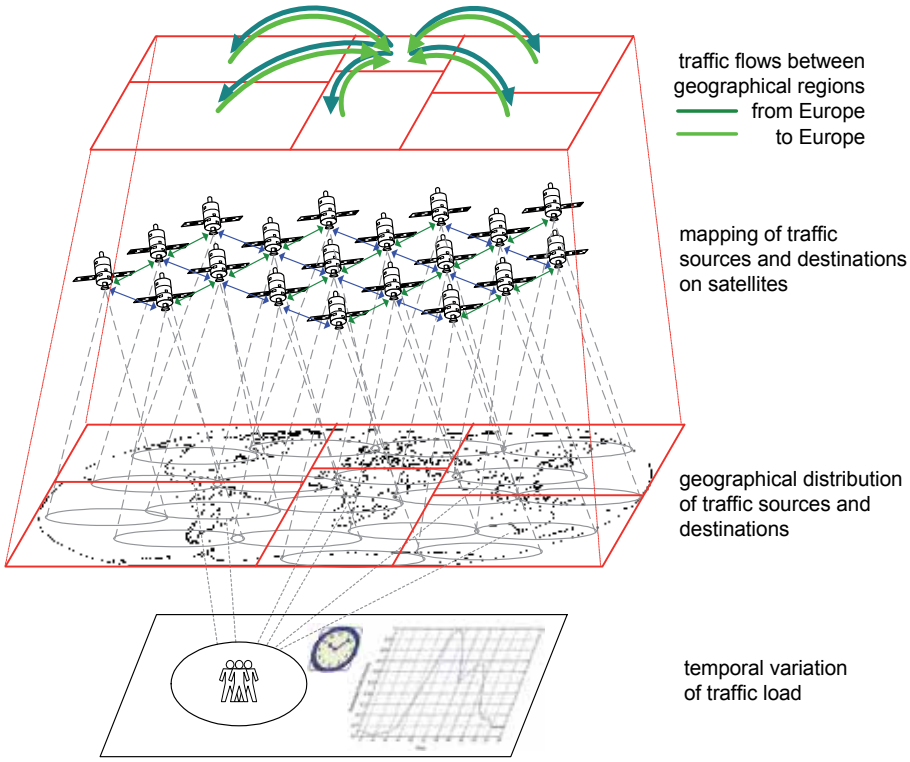


Fig. 7. Global aggregate traffic intensity model.

With the normalized cumulative traffic on each satellite, which is proportional to the intensity of traffic sources in the satellite's coverage area, it is possible to modulate the selected traffic source generator (not shown in Fig. 7). Thus data packets are actually generated considering the relative traffic intensity experienced by a particular satellite.

The destination satellite is selected for each packet in accordance with the traffic flow pattern. The probability of selecting a given satellite as a destination node is proportional to its coverage share in the destination region divided by the sum of all coverage shares in that region. Thus, although in a simplified manner, the model is taking into consideration also multiple coverage. In the case of using different traffic source models to generate distinct types of traffic by global aggregate traffic intensity model, one should also consider different, service specific traffic flow patterns.

4. Summary

Traffic engineering involves adapting the routing of traffic to the network conditions with two main goals: (i) providing sufficient quality of service, which is important from user's point of view, and (ii) efficient use of network resources, which is important for operators of telecommunication's network. The presented routing and traffic engineering issues addressed both goals that are explained using the ISL network as a concrete example of highly dynamic telecommunication network with several useful properties, which can be exploited by developing of routing procedures. However, the presented work is not limited to ISL networks, but can be used also in other networks as described in (Liu et al., 2011; Long et al., 2010; Rao & Wang, 2010, 2011). Routing and traffic engineering functions are presented in modular manner for easier reuse of particular procedures.

Adaptation of routing requires, in addition to good understanding of the fundamental network operating conditions, also good knowledge of the characteristics of different types of traffic in the network. In order to support better modelling of traffic characteristics a modular methodology is described for developing a global aggregate traffic intensity model suitable for supporting the dimensioning and computer simulations of various procedures in the global networks. It is based on the integration of modules describing traffic characteristics on four different levels of modelling, i.e. geographical distribution of traffic sources and destinations, temporal variations of traffic sources' intensity, traffic flows patterns and statistical behaviour of aggregated traffic sources.

5. References

Bertsekas D. & Gallager R. (1987). *Data Networks*, Englewood Cliffs: Prentice-Hall International.

Franck L. & Maral G. (2002a). Signaling for inter satellite link routing in broadband non GEO satellite systems. *Computer Networks*, Vol. 39, No. 1, pp. 79-92.

Franck L. & Maral G. (2002b). Routing in Networks of Intersatellite Links. *IEEE Transaction on Aerospace and Electronic Systems,* Vol. 38, No. 3, pp. 902-917.

Hu Y. F. & Sheriff R. E. (1997). The Potential Demand for the Satellite Component of the Universal Mobile Telecommunication System. *Electronics and Communication Engineering Journal*, April 1997, pp. 59-67.

Hu Y. F. & Sheriff R. E. (1999). Evaluation of the European Market for Satellite-UMTS Terminals. *International Journal of Satellite Communications*, Vol. 17, pp. 305-323.

Liu, X.; Ma, J. & Hao, X. (2011). Self-Adapting Routing for Two-Layered Satellite Networks. *China Communications*, Volume 8, Issue 4, July 2011, pp. 116-124.

Long F; Xiong N.; Vasilakos A.V.; Yang L.T. & Sun, F. (2010). A sustainable heuristic QoS routing algorithm for pervasive multi-layered satellite wireless networks. *Wireless Networks*, Volume 16, Issue 6, August 2010, Pages 1657-1673.

Micheel, 2002. National Laboratory for Applied Network Research), Passive Measurement and Analysis. http://pma.nlanr.net/PMA/, 22 October, 2002.

Mohorcic M.; Svigelj A. & Kandus G. 2004. Traffic Class Dependent Routing in ISL Networks. *IEEE Transaction on Aerospace and Electronic Systems,* Vol. 39, pp. 1160-1172.

Mohorcic M.; Svigelj A.; Kandus G. & Werner M. (2000). Comparison of Adaptive Routing Algorithms in ISL Networks Considering Various Traffic Scenarios. In: *Proc. of 4th*

European Workshop on Mobile and Personal Satellite Communications (EMPS 2000), pp. 72-81, London, UK; September 18, 2000.

Mohorcic M.; Svigelj A.; Kandus G. & Werner M. (2002a). Performance Evaluation of Adaptive Routing Algorithms in Packet Switched Intersatellite Link Networks. *International Journal of Satellite Communications,* Vol. 20, pp. 97-120.

Mohorcic M.; Svigelj A.; Kandus G.; Hu Y. F. & Sheriff R. E. (2003). Demographically weighted traffic flow models for adaptive routing in packet switched non-geostationary satellite meshed networks. *Computer Networks*, No. 43, pp. 113-131.

Mohorcic M.; Svigelj A.; Werner M. & Kandus G. (2001). Alternate link routing for traffic engineering in packet oriented ISL networks. *International Journal of Satellite Communications,* No. 19, pp. 463-480.

Mohorcic M.; Werner M.; Svigelj A. & Kandus G. (2002b). Adaptive Routing for Packet-Oriented Inter Satellite Link Networks: Performance in various Traffic Scenarios. *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, pp. 808-818.

Rao Y. & Wang R. (2011). Performance of QoS routing using genetic algorithm for Polar-orbit LEO satellite networks. *AEU - International Journal of Electronics and Communications,* Vol. 65 (6), pp. 530-538.

Rao, Y. & Wang, R. (2010). Agent-based load balancing routing for LEO satellite networks. *Computer Networks*, Volume 54, Issue 17, 3 December 2010, pp. 3187-3195.

Ryu B., (1999). Modeling and Simulation of Broadband Satellite Networks: Part II - Traffic Modeling, *IEEE Communication Magazine*, July 1999.

Salvador P.; Pacheco A. & Valadas R. (2004). Modeling IP traffic: joint characterization of packet arrivals and packet sizes using BMAPs. *Computer Networks*, No. 44, pp. 335-352.

Svigelj A.; Mohorcic M. & Kandus G. (2004b) Traffic class dependent routing in ISL networks with adaptive forwarding based on local link load information. *Space communications*, Vol. 19, pp. 158-170.

Svigelj A.; Mohorcic M.; Franck L. & Kandus G. (2012). Signalling Analysis for Traffic Class Dependent Routing in Packet Switched ISL Networks. To appear in: *Space communications*, Vol. 22:2.

Svigelj A.; Mohorcic M.; Kos A.; Pustisek M.; Kandus G. & Bester J. (2004a). Routing in ISL networks Considering Empirical IP Traffic. *IEEE Journal on Selected Areas in Communications*, Vol. 22, No. 2, pp. 261-272.

Wang Z & Crowcroft J.(1996). Quality-of-Service Routing for Supporting Multimedia Applications. *IEEE Journal on Selected Areas in Communications,* Vol. 14, No. 7, pp. 1228-1234.

Wang Z. (1999). On the complexity of quality of service routing. *Information Processing Letters*, Vol. 69, pp. 111-114.

Werner M. & Lutz E. (1998). Multiservice Traffic Model and Bandwidth Demand for Broadband Satellite Systems. In proceedings*: M. Ruggieri (Ed.), Mobile and Personal Satellite Communications 3*, pp. 235-253, Venice, Italy, November 1998.

Werner M. & Maral G. (1997). Traffic Flows and Dynamic Routing in LEO Intersatellite Link Networks. In: *Proc. IMSC '97*, pp. 283-288, Pasadena, California, USA, June 1997.

Werner M; Frings J.; Wauquiez F. & Maral G. (2001). Topological Design, Routing and Capacity Dimensioning for ISL Networks in Broadband LEO Satellite Systems. *International Journal of Satellite Communications*, No. 19, pp. 499-527.

Wood, L.; Clerget, A.; Andrikopoulos I.; Pavlou G. & Dabbous W. (2001). IP Routing Issues in Satellite Constellation Networks. *International Journal of Satellite Communications*, Vol. 19, No. 1, pp. 69 92.

15

Modeling and Simulating the Self-Similar Network Traffic in Simulation Tool

Matjaž Fras[1], Jože Mohorko[2] and Žarko Čučej[2]
[1]*Margento R&D, Maribor,*
[2]*University of Maribor, Faculty of Electrical Engineering and Computer Science, Maribor, Slovenia*

1. Introduction

Telecommunication networks are growing very fast. The user's needs, in regards to new services and applications that have a higher bandwidth requirement, are becoming bigger every day. A telecommunication network requires early design, planning, maintenance, continuous development and updating, as demand increases. In that respect we are forced to incessantly evaluate the telecommunication network's efficiency by utilizing methods such as measurement, analysis modeling and simulations of these networks.

Measuring, analyses and the modeling of self-similar traffic has still been one of the main research challenges. Several studies have been carried-out over the last fifteen years on: analysis of network traffic on the Internet [30], [31], traffic measurements in the high speed networks [32], and also measurement in the next generation networks [33]. Also, a lot of research works exist, where attention had been given to analysis of the network traffic caused by different applications, such as P2P [34], [35], network games [36] and VoIP application Skype [37]. Analyses of the measured network traffic help us to understand the basic behavior of network traffic. Various have showed that traffic in contemporary communication networks is well described with a self-similar statistical traffic model, which is based on fractal theory [6]. The pioneers in this field are: Leland, Willinger, and many others [1], [5], [6]. They introduced the new network traffic description in 1994. New description appeared as an alternative to traditional models, as were Poisson and Markov, which were used as a good approximation for telephone networks (PSNT networks) when describing the process of call durations and time between calls [5], [20]. These models do not allow descriptions of bursts, which are distinctive in today's network traffic. Such bursts can be described by a self-similarity model [5], [6], because it shows bursts over a wide-range of time scales. This contrasts with the traditional traffic model (Poisson model), which became very smooth during the aggregation process. The measure of bursts and also self-similarity present the Hurst parameter [1]-[4], which is correlated with another very important property called long-range dependence [5]-[8]. This property is also manifested with heavy-tailed probability of density distributions [5], [6], such as Pareto [43] or Weibull [44]. So Pareto's and Weibull's heavy-tailed distributions became the most frequently used distributions to describe self-similar network traffic in communication networks.

During past years another aspect of network traffic studying has also appeared. In this case, the network traffic is researched from application or data source point of view, especially focused on statistics of file sizes and inter-arrival times between files [19]. These research works are very important for describing a relation between packet network traffic on lower ISO/OSI layers and data source network traffic on higher layers of ISO/OSI model. Based on the research of WWW network traffic, it has been shown that file sizes of such traffic are best described by Pareto distribution with shape parameter $a = 1$ [38]. That was also shown for the FTP traffic, where the shape parameter of Pareto distribution is in the range $0.9 < a < 1.1$ [20]. In [6], [39], and [40] it is shown that inter-arrival time of TCP connections are self-similar processes, which can be described by Weibull heavily tailed distribution.

With expansion of simulation tools, which are used for simulation of communication networks, the knowledge about simulating the network traffic also becomes very important. One of the important tasks in simulations is also knowledge about modeling and simulating of network traffic. Network traffic is usually modeled in simulation tools from an application point of view [42], [45]. It is usually supposed that the file size statistics and file inter-arrival times are known [39], [40]. Such kinds of traffic models are supported by most commercial telecommunication simulation tools such as the OPNET Modeler [10], [11], [24], used in our simulations and experiments. Consequently, for using the measured data of packet traffic, when modeling file statistics, it is necessary to transform packets' statistics into files' statistics [9, 10]. This transformation contains opposite operations in relation to the fragmentation and encapsulation process. Extensive research and investigation about traffic sources in contemporary networks show that this approach requires an in-depth analysis of packet's traffic (which needs specialized, very powerful and consequently, expensive instruments). This approach, in the case of encrypted packets and non-standard application protocols, is not completely possible. In such cases, capture of entire packets is also necessary, which can be problematic in contemporary high-speed networks. Another approach estimates distribution parameters of file data sources from measured packets' network traffic. For such approach, we have developed and tested different methods [42], [45]. Estimated distribution parameters are used for modeling of the measured network traffic for simulation purposes. Through the use of these methods we want to minimize discrepancies between the measured and simulated traffic in regards to an average bit rate and bursts, which are characteristic of self-similar traffic.

2. Network traffic

2.1 Packet network traffic measuring

The measuring and analyzing of real network traffic provide us with a very important knowledge about computer network states. In analyzing process, we need statistical mathematical tools. These tools are crucial for accuracy of a derived mathematical model, described by stochastic parameters for packet size and inter-arrival time [9]. Using this simulation model, we want to acquire information about telecommunication network's performances for:

- improvement of the current network,
- bottleneck searching,
- building and development of new network devices and protocols,

- and for ensuring quality of service (QoS) for real-time streaming multimedia applications.

Using this information, network administrators can make the network more efficient.

The simplest tools that measure and capture the packets of network traffic are packet sniffers. Packet sniffers, also known as protocol or network analyzers, are tools that monitor and capture network traffic with all content of network traffic. We can use sniffers to obtain the main information about network traffic, such as packet size, inter-arrival time and the type and structure of IP protocol. Sniffers have become very important and indispensable tools for network administrators. Figure 1 shows traffic captured by a packet sniffer.

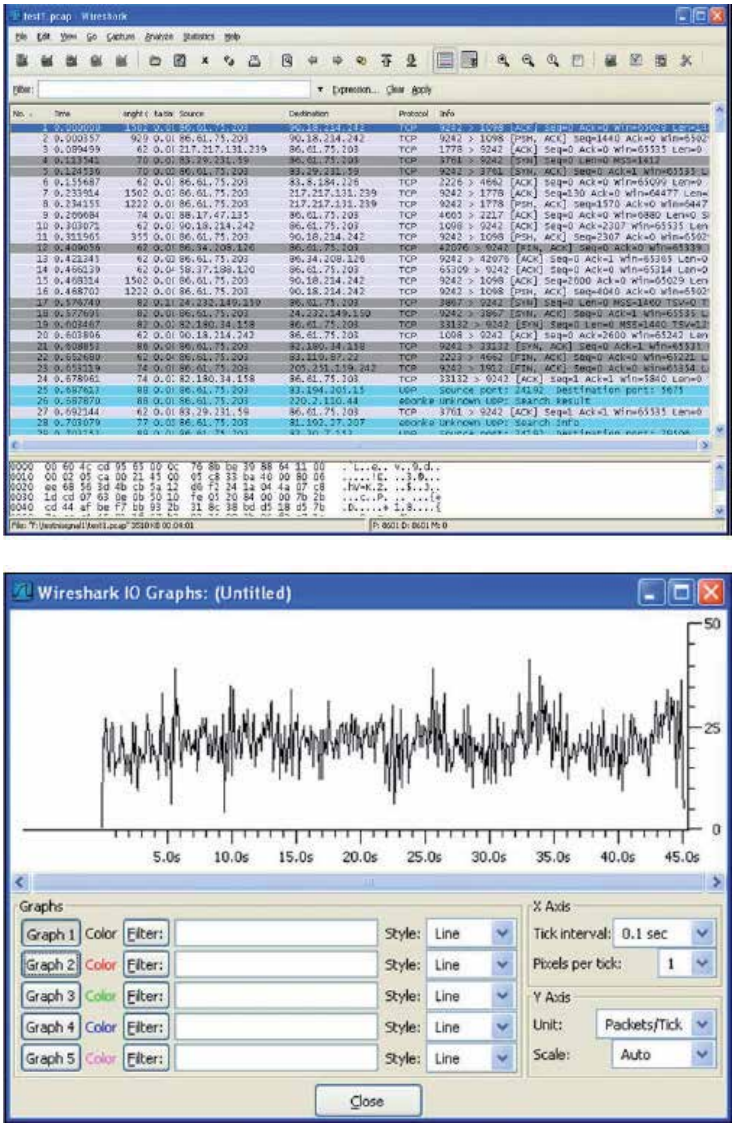


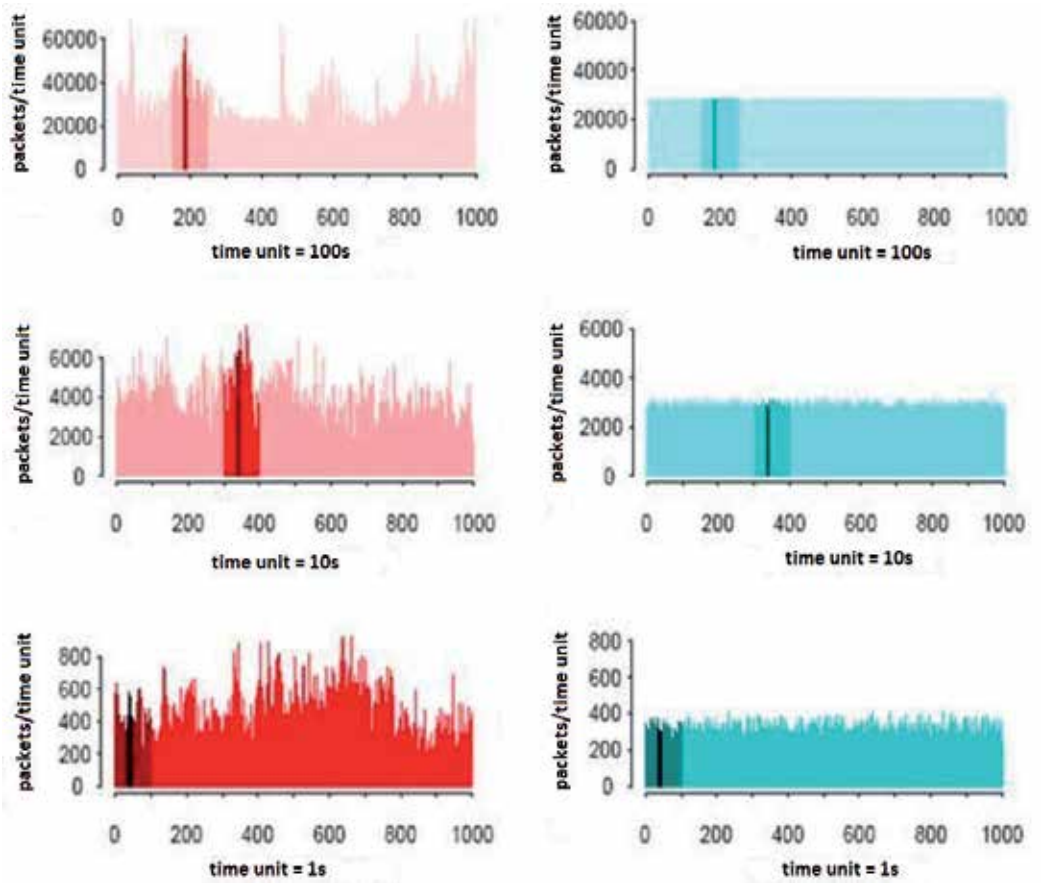
Fig. 1. User interface of WireShark sniffer during the network capturing.

Any sniffers are able to extract this data from the IP headers. Knowing them, it is then simple to calculate a length of IP PDU (Protocol Data Unit), which also contains a header of higher layer protocols. Using an in-depth header analysis, it is possible, in the similar way to the IP header, to calculate the lengths of all these headers.

An analytical description of network traffic does not exist, because we cannot predict the size and arrival time of the next packet. Therefore, we can only describe network traffic as a stochastic process. Hence, we have tried to describe these two stochastic processes (arrival time and packet size) with the use of Hurst parameter and probability distributions.

2.2 Self-similarity

In the 1990s, new descriptions and models of network's traffic were developed, which then replaced the traditional traffic models, such as Poisson and Markov [5], [20]. The Poisson process was widely used in the past, because it gave a good approximation of telephone network (PSNT networks), especially when describing times between each call and call durations. This model is usually described by exponential probability distribution, which is characterized by the parameter λ (number of events per second). However, these models do not allow for descriptions of bursts, which are distinctive in today's network traffic. Such



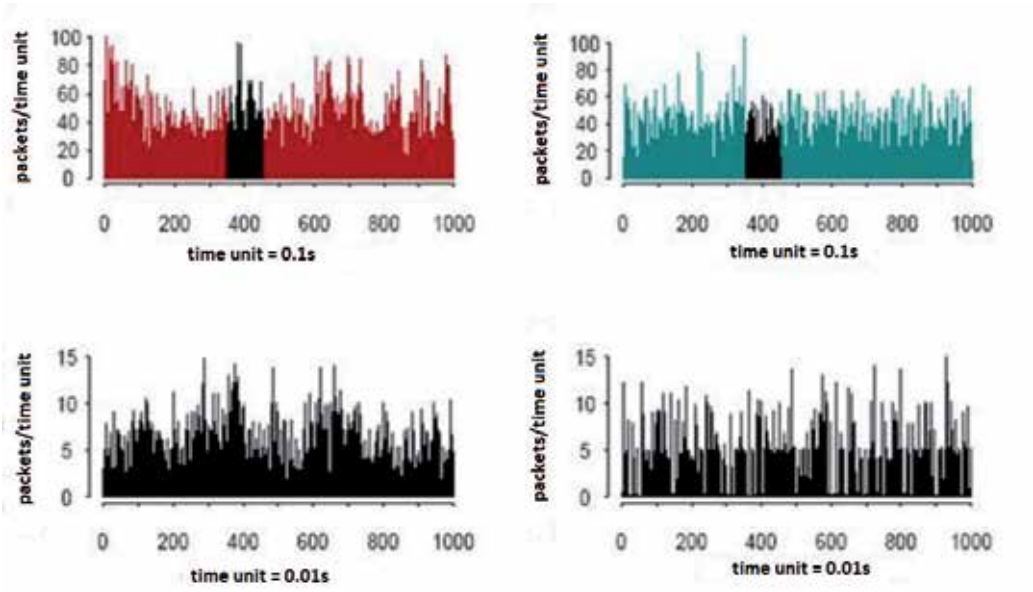


Fig. 2. Comparison of self-similar network traffic (left) and synthetic traffic created by Poisson model (right) on different time scales (100, 10, 1, 0.1 and 0.01s). Self similar traffic contains bursts on all time scales in contrast to the generated synthetic traffic, based on the Poisson model, which tends to average on longer time [1].

bursts can be described by a self-similarity model, because it shows bursts over a wide-range of time scales [1]-[4]. This contrasts the traditional traffic model (Poisson model), which becomes very smooth during the aggregation process.

2.3 Self-similarity

The definition of self-similarity is usually based on fractals for the standard stationary time series [5], [6], [21].

Let $X = (X_t, t = 0, 1, 2,\ldots)$ be a covariance stationary stochastic process; that is a process with a constant mean, finite variance $\sigma^2 = E[(X_t - \mu)2]$, with auto-covariance function $\gamma(k) = E[(X_t - \mu)(X_t+k - \mu)]$, that depends only on k. Then the autocorrelation function $r(k)$ is:

$$r(k) = \frac{\gamma(k)}{\sigma^2} = \frac{E\left[(X_t - \mu)(X_{t+k} - \mu)\right]}{E\left[(X_t - \mu)^2\right]}, \qquad k = 0,1,2,\ldots \tag{1}$$

Assume X has an autocorrelation function, which is asymptotically equal to:

$$r(k) \approx k^{-\beta} L_1(k), \quad k \to \infty, \quad 0 < \beta < 1, \tag{2}$$

where $L_1(k)$ slowly varies at infinity, that is $\lim_{t\to\infty}(L_1(tx) / L_1(t)) = 1$ for all x > 0. Such functions are for example $L_1(t) = const.$ and $L_1(t) = \log(t)$) [5], [6].

The measure of self-similarity is the Hurst parameter (H), which is in a relationship with the parameter β in equation (3).

$$H = 1 - \frac{\beta}{2} \tag{3}$$

Let's define the aggregation process for the time series [5], [6]:

For each m = 1, 2, 3, … let $X^{(m)} = (X_k^{(m)}, k = 1,2,..m)$ denote a new time series obtained by averaging the original series X over a non-overlapping block of size m. That is, for m=1, 2, 3, …, $X^{(m)}$ is given by:

$$X_k^{(m)} = \frac{1}{m}(X_{km-m+1} + ... + X_{km}), \qquad k = 1,2,3, ... \tag{4}$$

$X_k^{(m)}$ is the process with average mean and autocorrelation function $r^{(m)}(k)$ [6].

The process X is called an exactly second order with parameter H, which represents the measure of self-similarity if the corresponding aggregated $X^{(m)}$ has the same correlation structures as X and $\text{var}(X^{(m)}) = \sigma^2 m^{-\beta}$ for all m = 1, 2, … :

$$r^{(m)}(k) = r(k), \text{ for all } \quad m = 1,2, ... \quad k = 1,2, ... \tag{5}$$

The process X is called an asymptotically second order with parameter $H = 1 - \beta/2$, if for all k it is large enough,

$$r^{(m)}(k) \to r(k), \qquad m \to \infty \tag{6}$$

It follows from definitions that the process is the second order self-similar in the exact or asymptotical sense, if their corresponding aggregated process $X^{(m)}$ is the same as X or becomes indistinguishable from X-at least with respect to their autocorrelation function. The most striking property in both cases, exact and asymptotical self-similar processes, is that their aggregated processes $X^{(m)}$ possess a no degenerate correlation structure as $m \to \infty$. This contrasts with the Poisson stochastic models, where their aggregated processes tend to second order pure noise as $m \to \infty$:

$$r^{(m)}(k) \to 0, \quad m \to \infty, \quad k = 0,1,2,... \tag{7}$$

Network traffic with bursts is self-similar, if it shows bursts over many time scales, or it can be also said over a wide-range of time scales. This contrasts with traditional models such as Poisson and Markov, where their aggregation processes become very smooth.

2.4 Long-range dependence

The self-similar process can also contain a property of long-range dependence [5]-[8]. Long range dependence describes the memory effect, where a current value strongly depends upon the past values, of a stochastic process, and it is characterized by its autocorrelation function. This property has a stochastic process, which satisfies relation (2), order with relation $r(k) = \gamma(k)/\sigma^2$.

For $0 < H < 1$, $H \neq 1/2$ it holds [6]

$$r(k) \approx H(2H-1)k^{-2H-2}, \qquad r \to \infty \tag{8}$$

For values $0.5 < H < 1$ autocorrelation function $r(k)$ behavior, in an asymptotic mean, as $ck^{-\beta}$ for values $0 < \beta < 1$, where c is constant $c > 0$, $\beta = 2 - 2H$, and we have:

$$\sum_{k=-\infty}^{\infty} r(k) = \infty \,. \tag{9}$$

The autocorrelation function decays hyperbolically, as the k increases, which means that autocorrelation function is non-summable. This is opposite to the property of short-range dependence (SRD), where the autocorrelation function decays exponentially and the equation (9) has a finite value. Short and long-range dependence have a common relationship with the value of the Hurst parameter of the self-similar process [6], [21]:

- $0 < H < 0.5$ →SRD - Short Range Dependence
- $0.5 < H < 1$ →LRD - Long Range Dependence

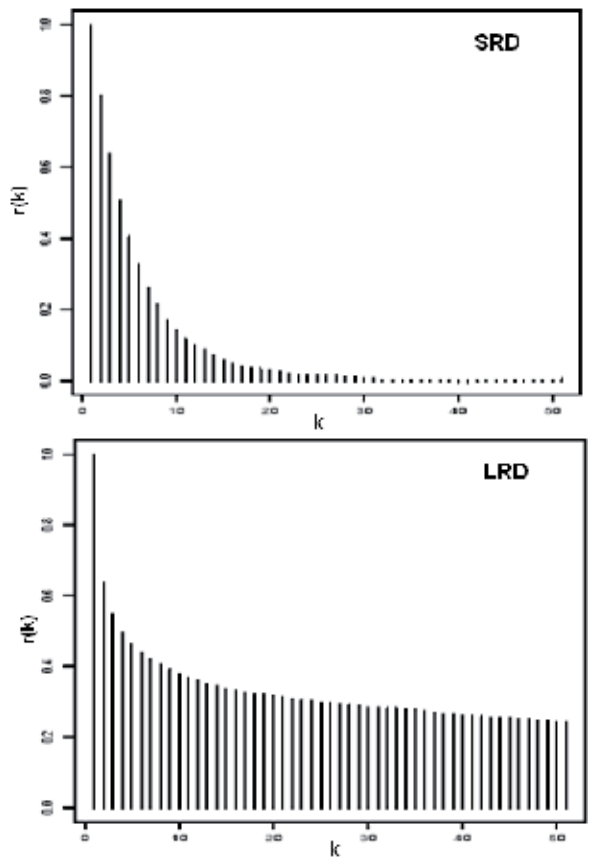


Fig. 3. Comparison between autocorrelation function of short range dependence process (left) and autocorrelation function of long range dependence process (right) [15].

2.5 Heavy-tailed distributions

Self-similar processes can be described by heavy-tailed distributions [5], [6], [9]. The main property of heavy-tailed distributions is that they decay hyperbolically, which is opposite to the light-tailed distribution, which decays exponentially. The simplest heavy-tailed distribution is Pareto. The probability density function of Pareto distribution is given by [43]:

$$p(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}, \quad k \le x, \quad \alpha, k > 0 \tag{10}$$

where parameter a represents the shape parameter, and k represents the local parameter of distribution (also a minimum possible positive value of the random variable x).

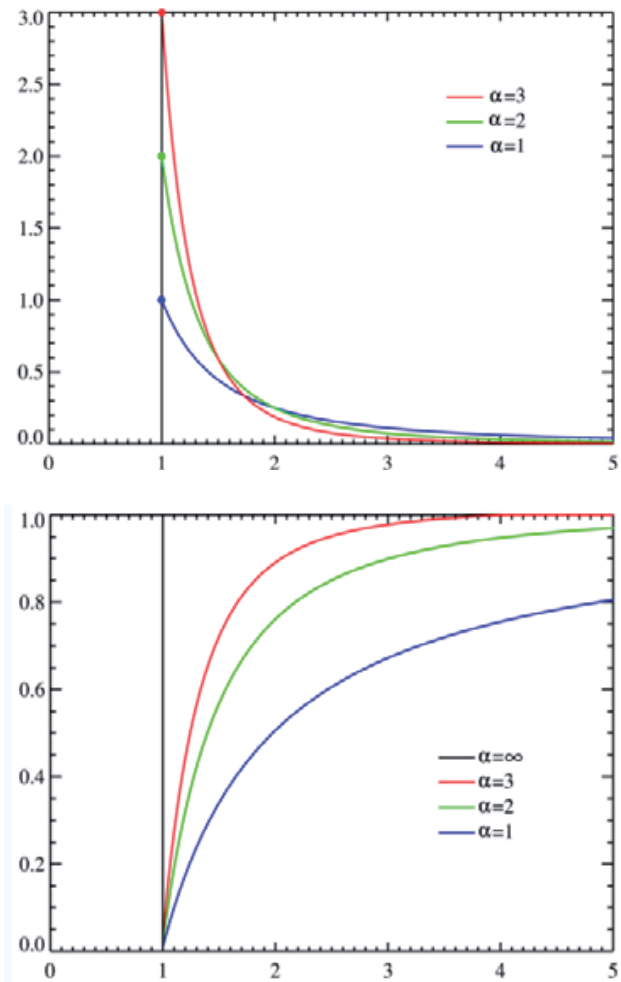


Fig. 4. Probability density function and cumulative distribution function of Pareto distribution for various shape parameters α and constant location parameter k = 1 [43].

Another very important heavy-tailed distribution is Weibull distribution, which is described by [44]:

$$p(x) = \frac{\alpha}{k} \cdot \left(\frac{x}{k}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{k}\right)^{\alpha}}, \quad x \geq 0, \quad \alpha, k > 0 \tag{11}$$

where parameter a presents the shape parameter, and k presents the local parameter of distribution.

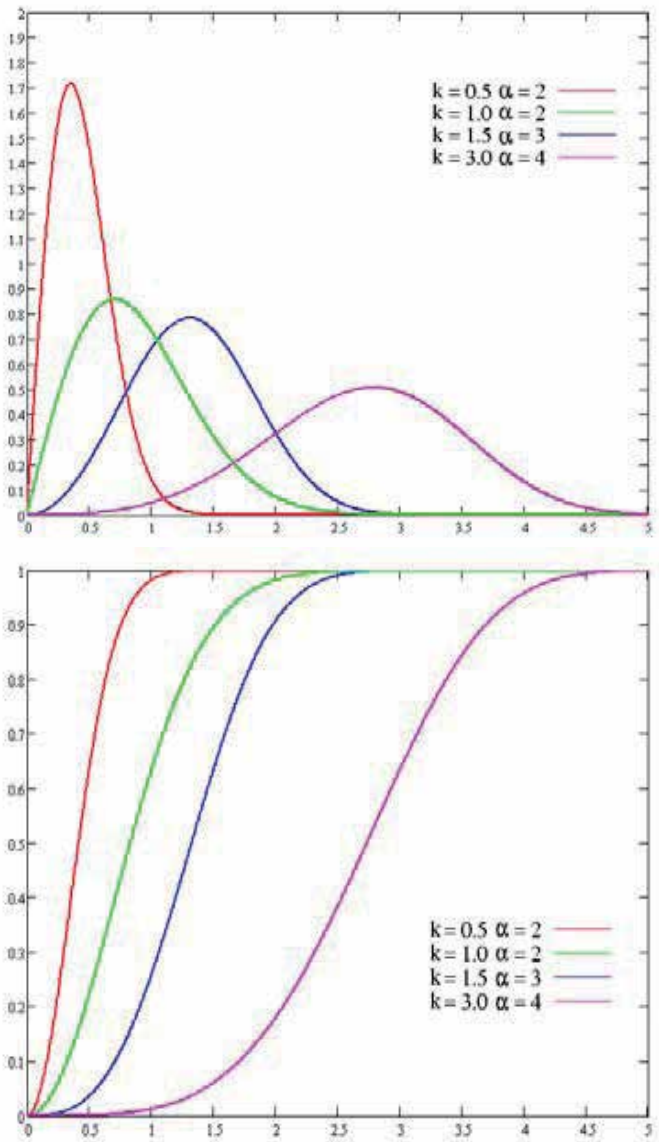


Fig. 5. Probability density function and cumulative distribution function of Weibull distribution for various shape parameters a and constant location parameter k [44].

2.6 Network traffic definitions

The network traffic can be observed on different layers of ISO/OSI model, for that reason we define different kinds of network traffics. The network traffic can be represented as a stochastic process, which can be interpreted as the traffic volume – measured in packets, bytes or bits per time unit, and it is consequent on data or packets, which are sent through the network in time unit. If we observe network traffic on the low level of ISO/OSI model, then define the packet network traffic [45] $Z_p[n]$:

Let define the packet network traffic $Z_p[n]$ as a stochastic process interpreted as the traffic volume, measured in packets per time unit. $Z_p[n]$ can be described as a composite of two stochastic processes:

$$Z_p[n] = X_p[n] \circ Y_p[n], \quad n \in \mathbb{R}. \tag{12}$$

where $X_p[n]$ represents packet size process and $Y_p[n]$ represents the packet inter-arrival time.

Packet-size process $X_p[n]$ is defined as a series of packet sizes l_{Pi} measured in bits (b) or bytes (B).

$$X_p[n] = \{l_{P1}, l_{P2}, ...l_{Pi}, ..., l_{Pn}\}, \quad 1 \le i \le n \tag{13}$$

where sizes of packets' l_{Pi} are limited by the shortest l_m and the longest l_{MTU} packet size (MTU - Maximum Transmission Unit).

$$l_m \le l_{Pi} \le l_{MTU} \tag{14}$$

Packet inter-arrival time process $Y_p[n]$ is defined as a series of times between packet arrivals t_{Pi} (time stamps).

$$\begin{aligned} Y_p[n] &= \{t_{P2} - t_{P1}, ..., t_{Pi} - t_{pi-i}, ..., t_{Pn} - t_{Pn-i}\}, \quad 1 \le i \le n \\ &= \{\Delta t_{p1}, \Delta t_{p2}, ..., \Delta t_{pi}, ..., \Delta t_{pn-1}\}, \quad 1 \le i \le n \end{aligned} \tag{15}$$

The measured network traffic is packet network traffic, which can be captured using special software program or hardware devices. For that reason, the measured network traffic is marked as $Z_{pm}[n]$. We also define modeled (simulated) network traffic as $Z_{ps}[n]$. We suppose, that the measured and modeled traffic is statistically equal, denoted by the symbol ≈,

$$Z_{pm}[n] \approx Z_{ps}[n] \tag{16}$$

if there are also statistical equalities between a packet size and inter-arrival time processes of measured, and modeled traffic.

$$X_{pm}[n] \approx X_{ps}[n]$$

and

$$Y_{pm}[n] \approx Y_{ps}[n] \tag{17}$$

Let's define network traffic on higher layers (application) of ISO/OSI model. Data source network traffic $Z_d[n]$ can be described as a composite of data source lengths $X_d[n]$ and data inter-arrival times $Y_d[n]$ processes:

$$Z_d[n] = X_d[n] \circ Y_d[n], \quad n \in \mathbb{R} \tag{18}$$

To provide statistical equality between packet network traffic $Z_p[n]$ and data sources network traffic $Z_d[n]$, we have performed a transformation between packet size process $X_p[n]$ and the process of data length $X_d[n]$ as well as transformation between packet inter-arrival time $Y_p[n]$ and data inter-arrival time $Y_d[n]$.

$$X_{pm}[n] \xrightleftharpoons{transformation} X_d[n] \tag{19}$$

$$Y_{pm}[n] \xrightleftharpoons{transformation} Y_d[n] \tag{20}$$

Transformation (19) and (20) allows estimation of packet traffic processes from data source traffic processes or vice verse.

3. Network traffic analysis and modeling

3.1 Hurst parameter estimations

Hurst's parameter represents the measure of self-similarity. There are several methods for estimating Hurst's parameter (H) [1]-[4] of stochastic self-similar processes. However, there are no criteria as to which method gives the best results. There are several different methods for estimating the Hurst parameter which can lead to diverse results [9], [10]. This is the reason why Hurst's parameter cannot be calculating but can be estimated. The most often used methods for Hurst's parameter estimation are [6], [8], [21]:

- Variance method is a graphical method, which is based on the property of slowly decaying variance. In a log-log scale plot, a sample variance versus a non-overlapping block of size m is drawn for each aggregation level. From the line with slope β we can estimate Hurst's parameter as a relationship, from equation (3).
- R/S method is also a graphical method. It is based on a range of partial sums regarding data series deviations from mean value, rescaled by its standard deviation. The slope in the log-log plot of the R/S statistic versus aggregated points is the estimation for Hurst's parameter.
- Periodogram method plots spectral density in a logarithm scale versus frequency (also in logarithm scale). The slope in periodogram allows the estimation of parameter H.

Figure 6 presents an example of test traffic and estimations of Hurst's parameter through different methods.

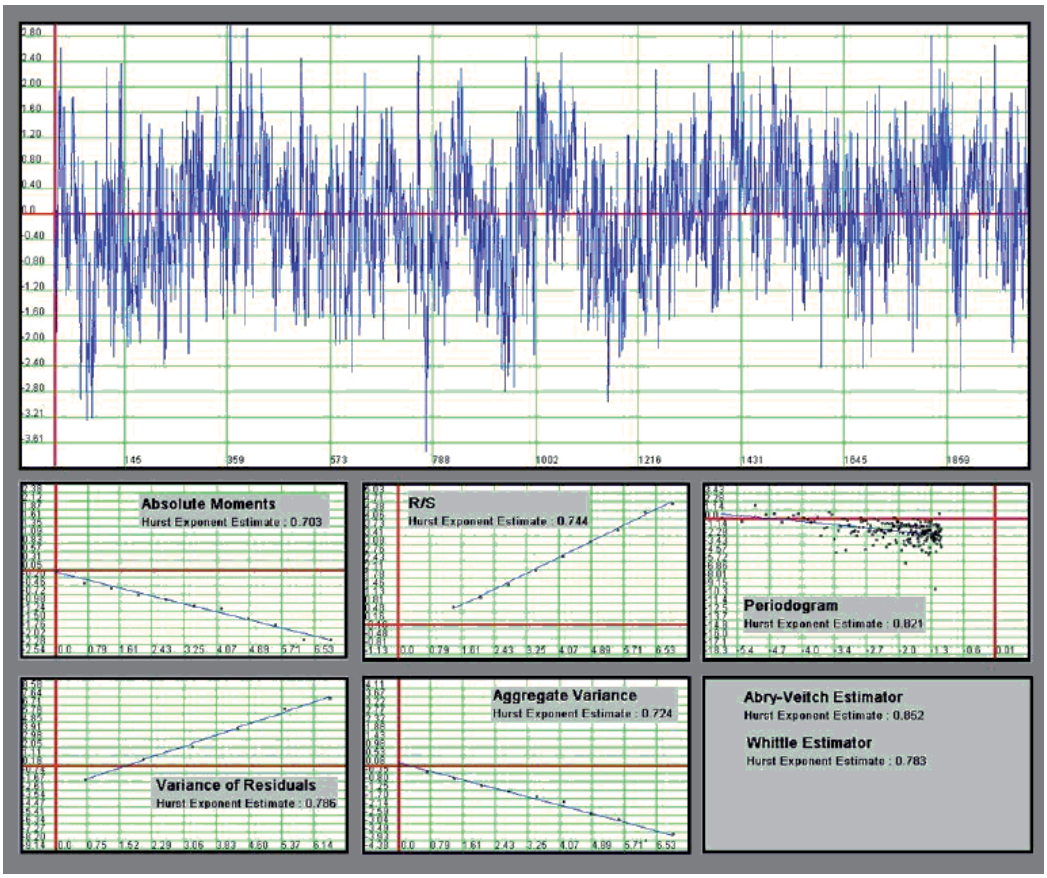


Fig. 6. Estimating parameter *H* for self-similar traffic (upper-left) with the variances method (lower left), R/S method (upper-right) and periodogram method (lower-right) using SELFIS tool [8].

3.2 Distribution parameter estimation for stochastic process of network traffic

Network traffic can be described by two stochastic processes, one for packet/data sizes and one for packet/data inter-arrival time. All processes are usually described by probability distributions. Self-similar process can be described by heavy tailed distributions. The main task for modeling the stochastic process with probability distribution is to choose the right distribution, which would be a good representation of our network traffic stochastic process. The statistic distribution parameters of data sources are then estimated by fitting tools [9], [25], [26] or other known methods, such as CCDF [6] or Hill estimator [17], [18]. Mathematical fitting tools are used (EasyFit), which allow us to automatically include the fit distribution of the stochastic process, and also estimate parameters of distribution from the captured traffic [9], [29].

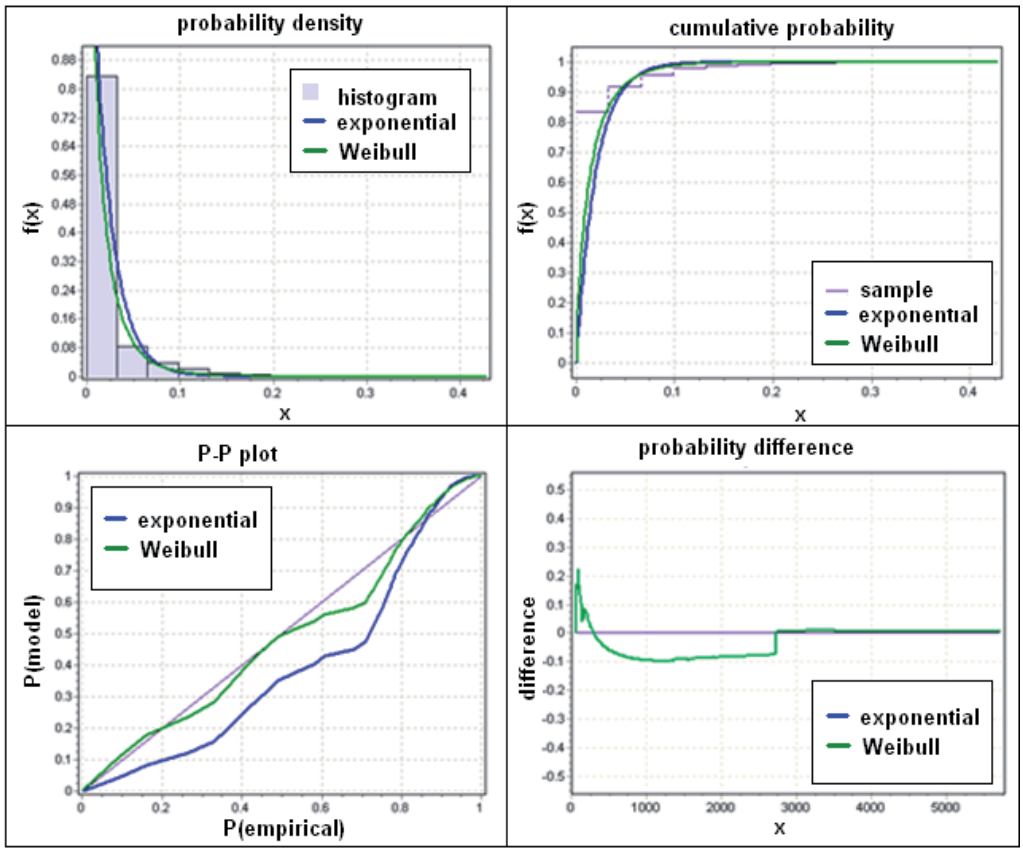


Fig. 7. For the stochastic process of inter-arrival time, distribution and estimate parameters of these distributions are chosen based on the histogram (upper left), and cumulative distribution function (upper right). Differences between empirical and theoretical distributions in P-P plot (lower left), and deferential distribution (lower right).

4. Simulation of network traffic in simulation tools

One of the very important tasks in simulation is modeling the real network parameters and network elements for simulation purposes. The main goal in successful modeling of network traffic is to minimize discrepancies between the measured simulations and by simulations statistically-modeled and generated traffic. This means, that both traffics are similar within the different criteria, such as bit and packet-rate, bursts (Hurst's parameter), variance, etc.

Network traffic simulations are usually based on modeling of data sources or applications. One of the most known simulation tools is OPNET Modeler [22], [23]. A simulation of network traffic in this tool is based on the "on/off" models [41] or more often used traffic generators. Difference between these manners is in a modeling manner. In the first case, the arrival process is described by Hurst's parameter (H) and the data length process is

described by probability density function (*pdf*). In the second case, processes of data length and data inter-arrival time are both described by *pdf*.

In OPNET Modeler, two standard node models appear [9]:

- Raw Packet Generator (RPG)
- IP station

Raw Packet Generator (RPG) is a traffic source model [16], [27] implemented specially to generate self-similar traffic, which is based on different fractal point processes (FPP) [41]. Self similar traffic is modeled with an arrival process, which is described by Hurst's parameter and the distribution probability for packet sizes. This arrival process can be based on many different parameters, such as Hurst parameter, average arrival rate, fractal onset time scale, source activity ratio and peak to mean ratio [16]. There are several different fractal point processes (FPP). In our case, we used the superposition of the fractal renewal process (Sub-FRP) model, which is defined as the superposition of M independent and probably identical renewal fractal processes. Each FRP stream is a point renewal processes and M numbers of independent sources compose the Sub-FRP model. Common inter-arrival probability density function $p(t)$ of this process is:

$$p(t) = \begin{cases} \gamma A^{-1} e^{-\gamma t / A} & 0 \le t \le A \\ \gamma e^{-\gamma} A^{\gamma} t^{-(\gamma+1)} & t \ge A \end{cases} \tag{21}$$

where $1 < \gamma < 2$. Process FRP can be defined as Sup-FRP process, when the number of independent identical renewal processes (M) is equal to 1. A model Sub-FRP is described by three parameters: γ, A and M. γ represents the fractal exponent, A is the location parameter, and M is the number of sources. These three parameters are in relationship with three OPNET parameters. These parameters are Hurst's average arrival-rate λ, and fractal onset time-scale (FOTS). The relationships between these three parameters of Sub-FRP and parameters in OPNET model are:

$$H = (3 - \gamma) / 2$$

$$\lambda = M\gamma[1 + (\gamma - 1)^{-1} e^{-\gamma}]^{-1} A^{-1} \tag{22}$$

$$T^{\alpha} = 2^{-1}\gamma^{-2} e^{-\gamma} (\gamma - 1)^{-1} (2 - \gamma)(3 - \gamma)[1 + (\gamma - 1)e^{\gamma}]^2 A^{\alpha},$$

where $\gamma = 2 - \beta$. Hurst parameter H is defined by equation (3). In the Sub-FRP model from OPNET, we can set Hurst's parameter (H), average arrival-rate (λ) and fractal onset time-scale (FOTS) in seconds. The recommended value for the parameter FOTS in OPNET is 1 second.

The IP station [16] can contain an arbitrary number of independent simultaneous working-traffic generators. Each generator enables the use of heavy-tailed distributions, such as Pareto or Weibull, for the generation of a self-similar network traffic by two distributions, one for length of a data source process and another for data inter-arrival time process. In our research, a traffic generator contained in an Ethernet IP station model of the OPNET Modeler simulation tool is used, as shown in the Figure 8.

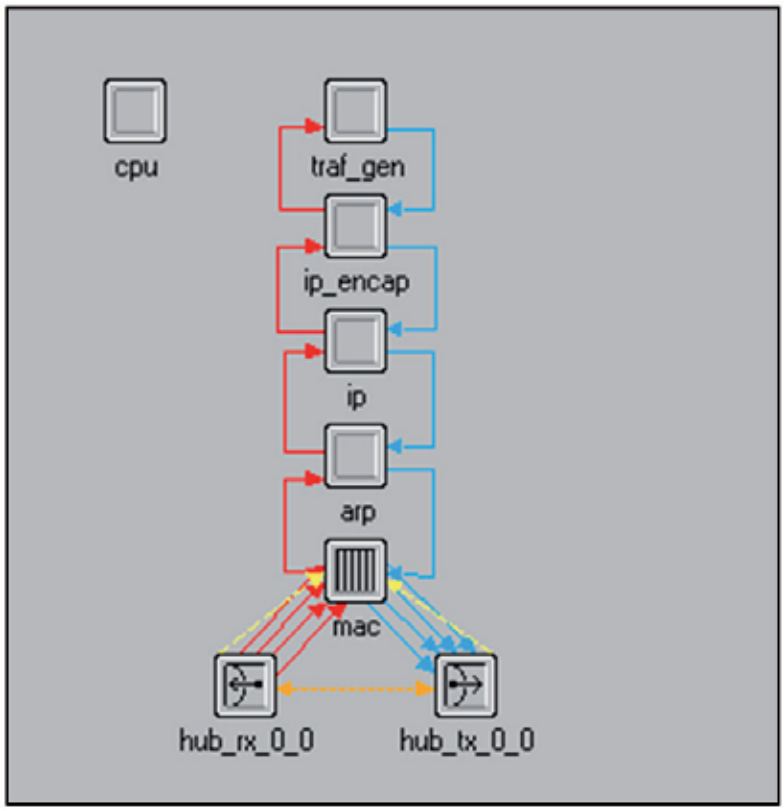


Fig. 8. Node model for used IP station in simulation.

In the IP station model, the traffic generator is placed above the IP encapsulation layer, which takes care of packets' formations and fragmentation. This is the process of segmentation of long data into the shorter packets, or vice versa, according to the RFC 793 [12]. Padding of the packet data payload with additional bits is also performed when data is shorter than a predefined minimal payload. Because the traffic is modeled, above IP level of the TCP/IP model, to the lengths of the generated data, 20 bytes of IP header are added. 18 bytes of information for MAC (14bytes) and CRC (4 bytes) are also further added. Structure of Ethernet frame used in the IP station model. Using this model, the applications' protocol does not impact the generated traffic. The model is suitable for the simulation cases, when we want to statistically model the network traffic, which can be caused by many arbitrary communications' applications. Using this approach, we can model such network traffic by single traffic source.

5. Estimation of simulation parameters of measured network traffic

The main problem of measured packet network traffic modeling is to estimate the parameter, which is needed for modeling measured network traffic in simulation tools. It has already been mentioned that the parameters of data source traffic processes are needed. We already described that transformation from packet network traffic $Z_p[n]$ to data source

network traffic $Z_d[n]$ is needed (section 2.6) [45]. There are many possibilities to make a transformation from $Z_p[n]$ to $Z_d[n]$, which allows estimation of parameters of data source network traffic processes. We investigated two algorithms [28]:

1. algorithm with an in-depth analysis of all packet headers,
2. algorithm with a coarse inspection of IP header only.

The main differences between them are complexity and the needed execution time. The first algorithm mimics a complete decapsulation process, and defragmentation in higher layers of the communication model. Any sniffers are able to extract this data from the IP header. Knowing them, it is then simple to calculate a length of IP PDU (Protocol Data Unit) which also contains a header of higher layer protocols. Through the use of an in-depth header analysis, it is possible, in the similar way as the IP header, to calculate the lengths of all these headers. Each packed IP header has four the so-called fragmentation fields that contain information about data fragmentation, which is shown on Figure 9.

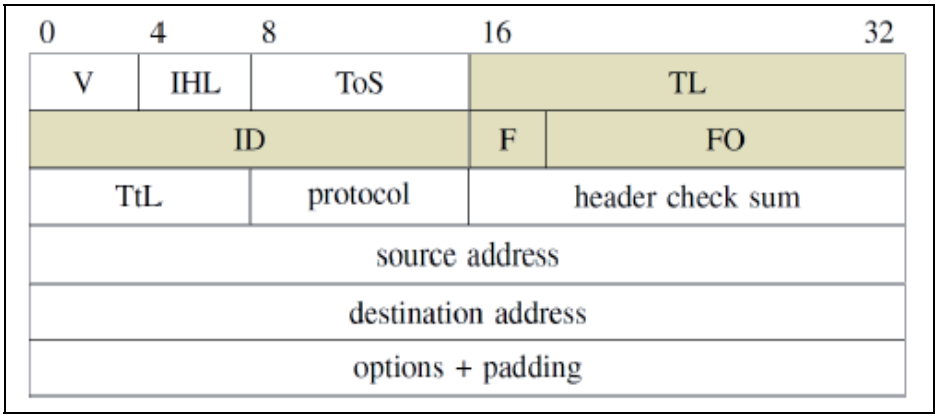


Fig. 9. IP header. Shadowed fields are used in the defragmentation process. Legend:
V: protocol version; IHL: Internet Header Length; ToS: Type of Service; TL: Total Length;
ID: Identification Data; F: Flags; FO: Fragment Offset; TTL: Time to Live.

Extensive research and investigation about traffic sources in contemporary networks show that this approach requires an in-depth analysis of packets (where need specialized, very powerful and consequently, expensive instruments), which in case of encrypted packets and non-standard application protocols, is not completely possible. In such cases, it is also necessary to capture the entire packets, which can be problematic in the high-speed networks. For these reasons, a simple algorithm has been developed, where only information of packets sizes, packet time stamps and IP addresses are needed.

The second algorithm skips decapsulation by considering the average lengths of packet headers and then uses only packet lengths and inter-arrival times. In the second case, the algorithm offers the estimation of data source network traffic, not the exact reconstructed data source traffic. The second algorithm represents the main part of method by mimic defragmentation process, which is described in detail in [45]. The main idea of mimic defragmentation process method is to compose data from the captured packet traffic, which is previously fragmented at the transmitter. The data source traffic estimation is

carried out by finding and summing fragmented packets' sequences without an in-depth analysis of packets. Fragmented sequence is defined as a sequence of l_{MTU} sized packets associated with the same source and destination addresses and terminated by packet shorter than l_{MTU}.

6. Simulation results

In real networks, we have captured packets of different network traffic through a Wireshark sniffer. The two different types of measured traffic are used for analysis, modeling and simulation purposes. These two test traffics are shown in Figure 10.

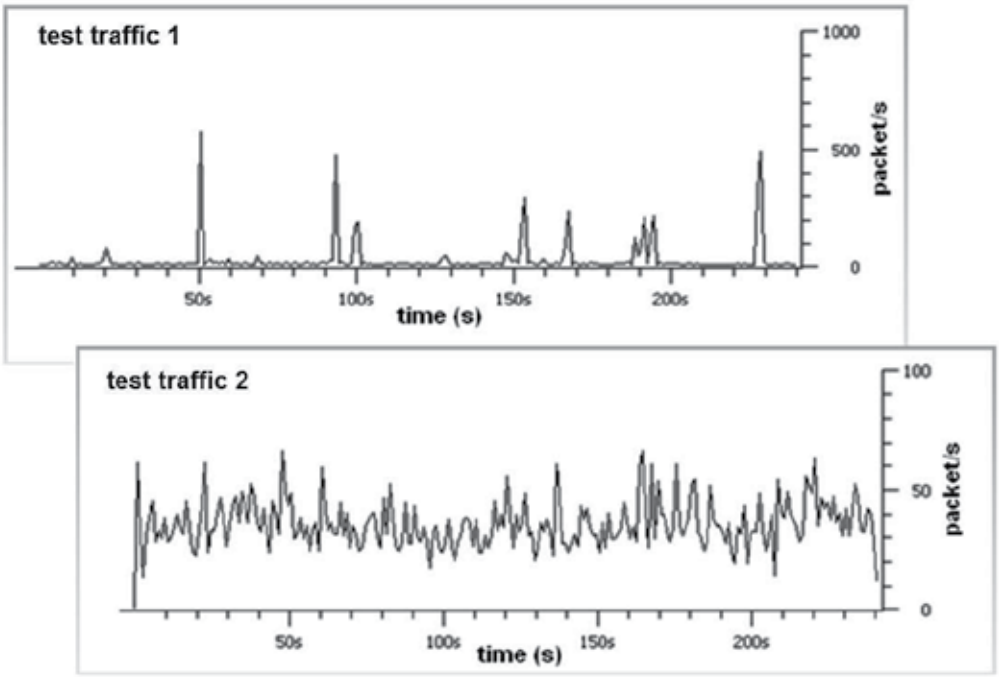


Fig. 10. Measured test traffic 1 and 2 captured by Wireshark sniffer.

measured test traffics	packet rate (p/s)	bit rate (kb/s)	variance method	R/S method	periodogram method
test traffic 1	24.02	108.90	0.630	0.723	0.843
test traffic 2	35.612	114.51	0.592	0.580	0.477

Table 1. The main properties of captured traffics. On the right side of the table the Hurst parameter is estimated using different methods for both test traffics.

For each of test traffics, the Hurst parameter has been estimated through different methods. The Hurst parameters for both cases are bigger than 0.5, so we can classify these test traffics

as a self-similar network traffic. Table 1 contains the estimated parameters *H* for both traffics, which are estimated by variance, R/S and periodogram methods. We also conducted tests about short and long-range dependence. In the case of the first test traffic, the autocorrelation function decayed hyperbolically, which means, that this traffic can have the property of a long-range dependence. For the second test traffic autocorrelation, function decayed exponentially towards 0. For this case, the sum of autocorrelations has finite results and, therefore, the test traffic 2 has the property of short-range dependence.

For both test traffics (test traffic 1 and test traffic 2) we estimate distribution and its parameters for data source traffic processes for simulation purpose. For that reason, we made an estimation of data source traffic from the captured packet traffic through the mimic defragmentation process method [45]. For both test traffics, the suitably heavy (Pareto or Weibull) and also light-tailed (exponential) distributions are chosen.

Based on the estimated distribution parameters for both measured test traffic (test traffic 1 and test traffic 2), we generated self-similar traffic in the OPNET simulation tool with two different station types – RPG and IP stations. We have created six different scenarios for each of test traffic. In the first two scenarios, the network traffic is generated by an RPG station, where a self-similarity is described by Hurst parameter. During the first scenario, we use heavy-tailed distribution for the data size process, while in the second a light-tailed distribution (exponential) is used. In the next four scenarios, network traffic is generated using the IP station, where we use different combination's distributions for the data size process and data inter-arrival time. One of the criterions, for successful modeling, is the difference between bit and packet-rates of the test traffic and modeled traffic in OPNET simulation tool. Besides the average values of bit and packet-rates, the more important criteria are also bursts' intensity within the network traffic. For each of test traffics (test traffic 1 and test traffic 2), the traffic which best represents the measured test traffic is chosen from six modeled traffics.

Test traffic 1 poses the property of long-range dependence, so there are a lot of bursts in the traffic. We model this measured-test traffic over six different scenarios. The results are shown in Figure 6 and Table 2. Table 2 shows the main properties of measured test traffic 1 and estimated distribution parameters which were used in OPNET simulation tool for simulating network traffic (the left side of Table 2). Table 2 (the right side) also shows main properties of simulated network traffics (six different scenarios) in OPNET simulation tool based on estimated distributions.

Table 2 shows modeling results for test traffic 1 over six different scenarios in OPNET simulation tool. There are estimated statistical parameters such as Hurst parameters and distributions used in models and simulation results using these models. Figure 11 shows all six modeled traffic traffics generated by OPNET, with estimated distributions and parameters from Table 2.

The best approximation for test traffic 1 is modeled traffic 5 from Table 2, which is described by Pareto distribution for data size process and Weibull distribution for data inter-arrival time. Figure 12 shows a comparison between the second test traffic and the modeled traffic 5 for bit rates. From all critera after comparison, we can say that the modeled traffic 5 is a good approximation of measured test traffic 1.

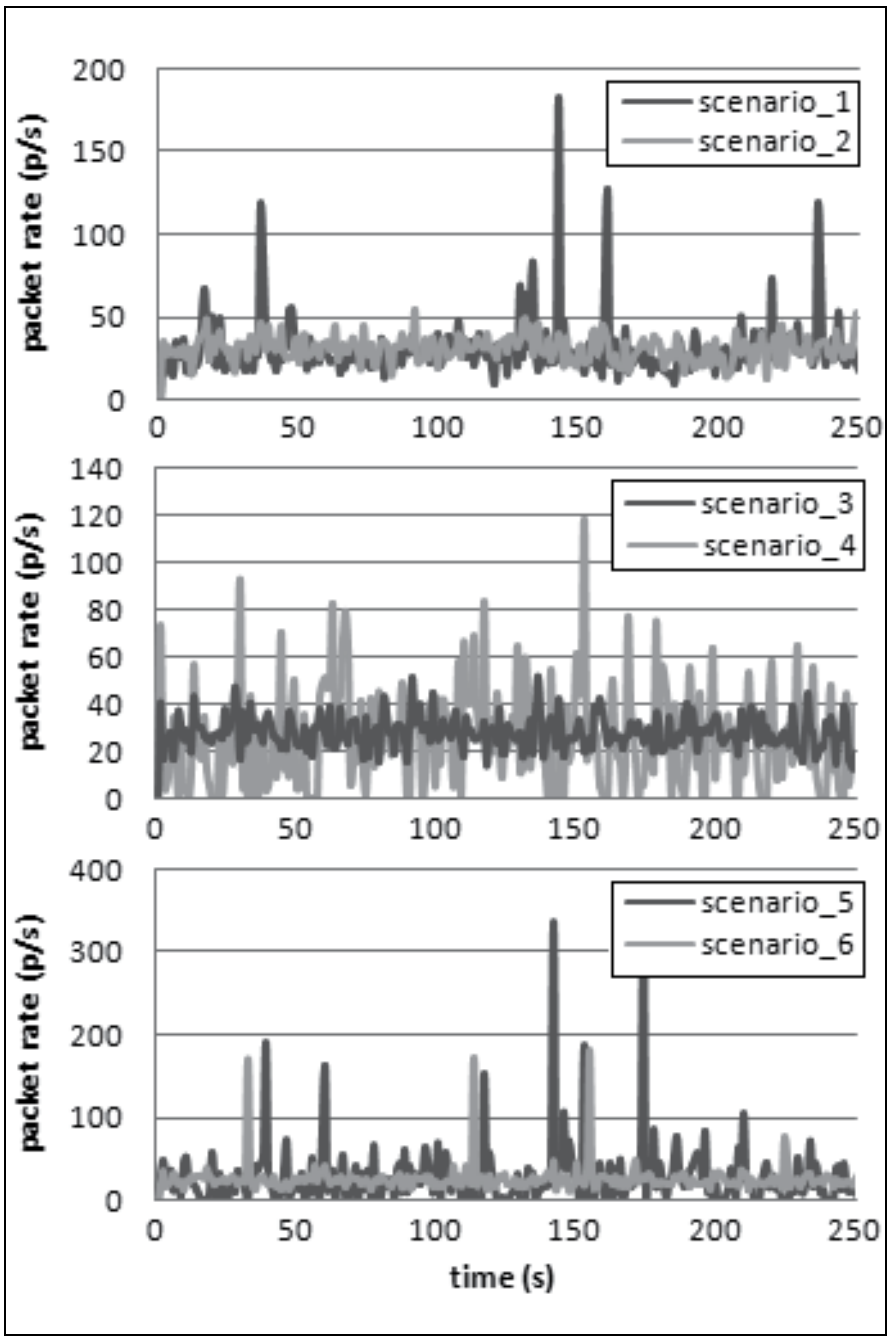


Fig. 11. Modeling measured test traffic 1 in OPNET simulation tool with six different estimated parameters from Table 2 (scenario 1 and 2 with RPG station, scenario 3, 4, 5, 6 with IP station).

parameters for modeling			parameters of measured and modeled traffic in OPNET		
traffic	data inter-arrival process	data size process	packet rate (p/s)	bite rate (kb/s)	H
measured test traffic 1	X	X	24	108.90	0.73
modeled 1	H = 0.732	Pareto a = 0.9835 β = 432	33.82	128.75	0.59
modeled 2	H = 0.732	exponential λ = 7547.2	29.18	181.44	0.59
modeled 3	exponential λ = 0.0458	exponential λ = 933.4	27.56	168.94	0.51
modeled 4	Weibull a = 0.304 β = 0.00578	exponential λ = 933.4	25.14	153.71	0.62
modeled 5	Weibull a = 0.304 β = 0.00578	Pareto a = 0.9835 β = 34	25.32	88.70	0.66
modeled 6	exponential λ = 0.0458	Pareto a = 0.9835 β = 34	26.63	81.30	0.55

Table 2. The left side of table shows the estimated distributions and parameters for measured test traffic 1 (six different distribution combinations). The right side of table shows main properties of modeled network traffic in OPNET simulation tool (six scenarios), where estimated distributions were used.

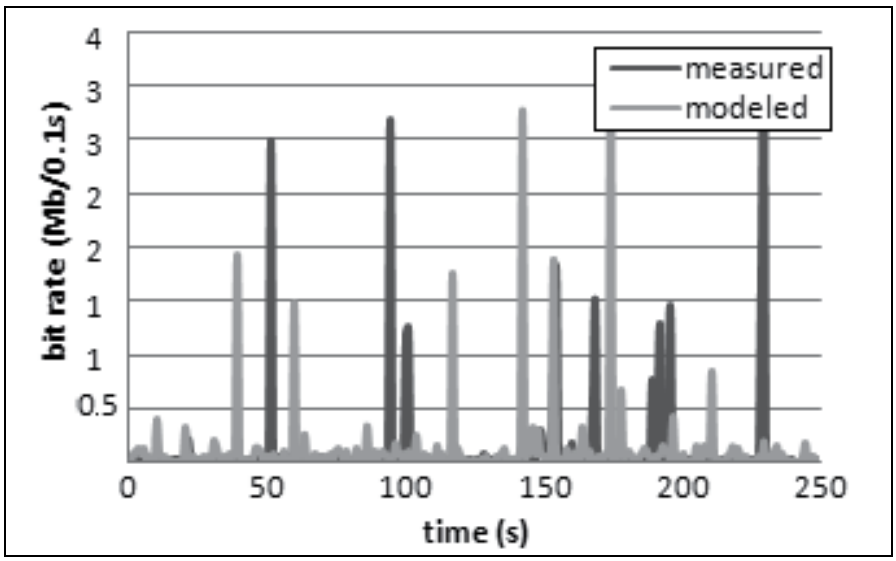


Fig. 12. Comparison between the modeled traffic 5 generated in OPNET simulation tool and the measured test traffic 1 in bits per second (kb/s).

Test traffic 2 is also modeled over six different scenarios, just like in the first case. Table 3 shows the main properties of measured test traffic 2 and estimated distribution parameters which were used in OPNET simulation tool for simulating network traffic (left side of Table 3). Table 3 (right side) also shows main properties of simulated network traffics (six different scenarios) in OPNET simulation tool.

As the best modeled traffic of test traffic 2 from all six cases (Table 3), we choose the case where simulated traffic is described by the exponential distribution for packet sizes and Weibull heavy-tailed distribution for inter-arrival time (modeled traffic 4). The bit-rate of this traffic is 33.27 (p/s) and packet-rate is 126.79 (kb/s), which are very close to the measured values. The Hurst parameter of the simulated traffic is 0.58, which is also close to the estimated values of the measured traffic. Figure 13 shows the comparison between the measured test traffic 2 and the best-modeled traffic (modeled traffic 4) for bit rates. From all critera after comparison, we can say that the simulated traffic is a good approximation of the measured traffic 2.

parameters for modeling			parameters of measured and modeled traffic in OPNET		
traffic	data inter-arrival process	data size process	packet rate (p/s)	bite rate (kb/s)	H
measured test traffic 2	X	X	35.61	114.51	0.55
modeled 1	$H = 0.55$	Pareto $a = 0.8373$ $\beta = 272$	49.46	231.98	0.62
modeled 2	$H = 0.55$	exponential $\lambda = 3619$	36.66	140.72	0.58
modeled 3	exponential $\lambda = 0.029$	exponential $\lambda = 452.48$	35.66	135.89	0.53
modeled 4	Weibull $a = 0.57$ $\beta = 0.01894$	exponential $\lambda = 452.48$	33.27	126.79	0.58
modeled 5	Weibull $a = 0.57$ $\beta = 0.01894$	Pareto $a = 0.8373$ $\beta = 34$	52.27	298.25	0.62
modeled 6	exponential $\lambda = 0.029$	Pareto $a = 0.8373$ $\beta = 34$	55.12	315.61	0.53

Table 3. The left side of table shows the estimated distributions and parameters for measured test traffic 2 (six different distribution combinations). The right side of table shows main properties of modeled network traffic in OPNET simulation tool (six scenarios), where estimated distributions and its parameters were used.

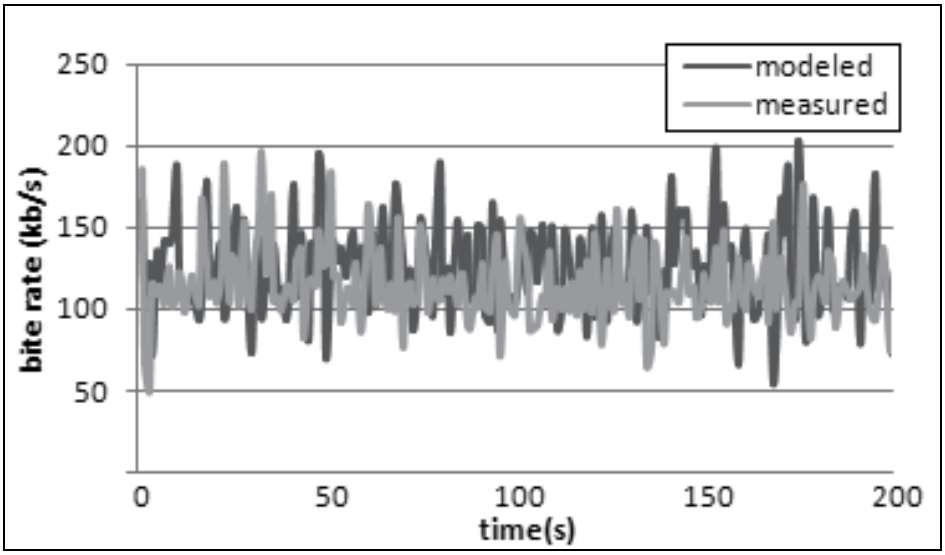


Fig. 13. Comparison between modeled traffic 4 generated in OPNET simulation tool and measured test traffic 2 in bits per second (kb/s).

7. Conclusion

In this chapter, we present our research in the area of measurements, modeling and simulations of the self-similar network traffic. Firstly, the state of the art method for modeling and simulating of self-similar network traffic is presented. We also describe a number of facts about self-similarity, long range dependences and probability, which are used to describe such stochastic processes. Described as well are the mechanism and models to simulate network traffic in the OPNET Modeler simulation tool. The main goal of our research is to simulate measured network traffic, where we tend to minimize discrepancies between the measured and the simulated network traffic in the sense of packet-rate, bit-rate, bursts intensity, and variances. One of the big challenges in our research work was to find appropriate method to estimate parameters of data source network traffic processes that are based on measured network packet's traffic. The estimated parameters are needed during the modeling of the measured network traffic in the simulation tool. For those reasons, we have developed different methods, which allow estimation of the parameters of data source network traffic processes, based on the measured network packet's traffic.

At the end of the chapter, all phases needed for simulating the measured network traffic in the OPNET simulation tool are presented. During the analysis phase we pay attention to the self-similar property, which has become the basic model for describing today's network traffic. In the network traffic theory, the properties of short and long-range dependence are directly prescribed by the values of estimated parameter *H*. In our network traffic analysis, we prove that network traffic (test traffic 2) can exist where Hurst parameter is bigger than 0.5, but this process does not have the property of a long-range dependence.

For the purpose of parameters estimation of data source network traffic processes, we have used a method that mimics packet defragmentation. Through the use of this method we

offer estimated parameters, used in simulations, where six traffics are simulated by different distributions for each of the measured test traffic. It can be seen from simulations that in the case of modeling self-similar traffic, short-range dependence is more appropriate for choosing exponential distribution to describe a packet-size process. The exponential distribution does not impact the extreme peaks in the modeled traffic. Pareto distribution is unsuitable for this purpose.

Heavy-tailed distributions, especially Pareto, are suitable for modeling a packet-size process of the measured network traffic, which are self-similar and also have the property of a long-range dependence (test traffic 1).

There are discrepancies between the measured and the modeled traffics in the sense of packet-rate, bit-rate, bursts intensity, and variances. With a method which mimics defragmentation, a good approximation of the measured network traffic is obtained. We cannot claim that this is the optimal method for all situations, because there are some limitations, although it shows good results through simulation in OPNET Modeler. We have noticed that estimating the shape-parameter of Pareto is very delicate, because a small deviation in the parameter causes large discrepancies regarding the network traffic's average values, which is one of the important criteria for traffic modeling.

8. Acknowledgment

This work has been partly financed by the Slovenian Ministry of Defense as part of the target research program "Science for Peace and Security": M2-0140 - Modeling of Command and Control information systems, and partly by the Slovenian Ministry of Higher Education and Science, research program P2-0065 "Telematics".

9. References

[1] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, On the self-similar nature of Ethernet traffic (Extended version), IEEE/ACM Transactions on Networking, Vol.2, pp.1-15, 1994.

[2] W. Willinger and V. Paxson, Where mathematics meets the Internet, Notices of the American Mathematical Society, 45(8): 961–970, 1998.

[3] K. Park, G. Kim and M. E. Crovella, On the Relationship Between File Sizes Transport Protocols, and Self-Similar Network Traffic, International Conference on Network Protocols, 171–180, Oct 1996.

[4] M. E. Crovella and A. Bestavros, Self-Similarity in World Wide Web Traffic Evidence and Possible Causes, IEEE/ACM Transactions on Networking, 1997.

[5] O. Sheluhin, S. Smolskiy and A. Osin, Self-Similar Processes in Telecommunications, John Wiley & Sons, 2007.

[6] K. Park and W. Willinger, Self-Similar Network Traffic and Performance Evaluation, John Wiley & Sons, 2000.

[7] T. Karagiannis, M. Molle and M. Faloutos, Understanding the limitations of estimation methods for long-range dependence, University of California.

[8] T. Karagiannis and M. Faloutos, Selfis: A tool for self-similarity and long range dependence analysis, University of California.

[9] M. Fras, J. Mohorko and Ž. Čucej, Estimating the parameters of measured self similar traffic for modeling in OPNET, IWSSIP Conference, 27.-30 June 2007, Maribor, Slovenia.
[10] J. Mohorko and M. Fras, Modeling of IRIS Replication Mechanism in a Tactical Communication network, using OPNET, Computer Networks, v 53, n 7, p 1125-36, 13 May 2009.
[11] J. Mohorko, M. Fras and Ž. Čucej: Modeling of IRIS replication mechanism in tactical communication network with OPNET, OPNETWORK 2007 - the eleventh annual OPNET technology Conference, August 27th-31st, Washington, D.C., 2007.
[12] RFC 793 - Transmission Control Protocol:. [Online]. Available: http://www.faqs.org/rfcs/rfc793.html
[13] M. Chakravarti, R. G. Laha and J. Roy, Handbook of Methods of Applied Statistics, Volume I, John Wiley and Sons, pp. 392-394, 1967.
[14] W. T. Eadie, D. Drijard, F. E. James, M. Roos and B. Sadoulet, Statistical Methods in Experimental Physics, Amsterdam, North-Holland, 269-271, 1971.
[15] A. Adas, Traffic Models in Broadband Telecommunication Networks, Communications Magazine, IEEE , vol 35/7, 82–89, 1997.
[16] J. Potemans, B. Van den Broeck, Y. Guan, J. Theunis, E. Van Lil and A. Van de Capelle, Implementation of an Advanced Traffic Model in OPNET Modeler, OPNETWORK 2003, Washington D.C., USA, 2003.
[17] B. Hill, A Simple Approach to Inference About tbc Tail of a Distribution, Annals of Statistics, Vol. 3, No. 5, 1975, pp.1163-1174.
[18] J. Judge, H. W. Beadle and J. Chicharo, Sampling HTTP response packets for prediction of web traffic volume statistics, IEEE Global Communications Conference (GLOBECOM'98), Sydney, Australia, Nov. 8-12, 1998.
[19] K. Park, G. Kim and M. E. Crovella, On the Relationship Between File Sizes Transport Protocols, and Self-Similar Network Traffic, International Conference on Network Protocols, 171–180, Oct 1996.
[20] V. Paxon and S. Floyd, Wide area traffic: the failure of Poisson modeling, IEEE/ACM Transactions on Networking, 3(3): 226–244, 1995.
[21] H. Yõlmaz, IP over DVB: Managment of self-similarity, Master of Science, Boğaziçi University, 2002.
[22] B. Vujičić, Modeling and Characterization of Traffic in Public Safety Wireless Networks, Master of Applied science, Simon Fraser University, Vancouver, 2006.
[23] M. Jiang, S. Hardy in Lj. Trajkovic, Simulating CDPD networks using OPNET, OPNETWORK 2000, Washington D.C., August 2000.
[24] J. Mohorko, M. Fras and Ž. Čučej, Modeling methods in OPNET simulations of tactical command and control information systems, IWSSIP Conference, 27.-30 June 2007, Maribor, Slovenia.
[25] A. M. Law and M. G. McComas, How the Expertfit distribution fitting software can make simulation models more valid, Proceedings of the 2001 Winter Simulation Conference.
[26] Free (demo) fitting tool EasyFit software [Online]. Available: www.mathwave.com/.

[27] F. Xue and S. J. Ben Yoo, On the Generation and Shaping Self-similar Traffic in Optical Packet-switched Networks, OPNETWORK 2002, Washington D.C., USA, 2002.
[28] Ž. Čučej and M.Fras, Data source statistics modeling based on measured packet traffic : a case study of protocol algorithm and analytical transformation approach, TELSIKS 2009, 9th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, Serbia, Niš, 7-9 October, 2009.
[29] M. Fras, J. Mohorko and Ž. Čučej, Analysis, modeling and simulation of P2P file sharing traffic impact on networks' performances. Inf. MIDEM, 38(2):117–123, 2008.
[30] H. Abrahamsson, Traffic measurement and analysis, Swedish Institute of Computer Science, 1999.
[31] C. Williamson, Internet traffic measurement, IEEE internet computing, vol. 5, no. 6, pp. 70–74, 2001.
[32] P. Celeda, High-speed network traffic acquisition for agent systems, in Proc. IEEE/WIC/ACM International Conference on High-Speed Network Traffic Acquisition for Agent Systems, Intelligent Agent Technology, November 2-5, 2007, pp. 477–480.
[33] D. Pezaros, Network Traffic Measurement for the Next Generation Internet. Computing Department Lancaster University, 2005.
[34] D. Epema, J. Pouwelse, P. Garbacki and H. Sips, The bittorrent P2P filesharing system: Measurements and analysis. Peer-to-Peer Systems IV, 2005.
[35] S. Saroiu, P. K. Gummadi and S. D. Gribble, A Measurement Study of Peer-to-Peer File Sharing Systems, in Proc. of the Multimedia Computing and Networking (MMCN), January 2-5, San Jose, Ca, USA, 2002.
[36] E. Asensio, J. M. Orduna and P. Morillo, Analyzing the Network Traffic Requirements of Multiplayer Online Games, in Proc. 2nd International Conference on Advanced Engineering Computing and Applications in Sciences: ADVCOMP'08, 2008, pp. 229–234.
[37] Y. Yu, D. Liu, J. Li and C. Shen, Traffic Identification and Overlay Measurement of Skype, in Proc. International Conference on Computational Intelligence and Security, November 3-6, vol. 2, 2006, p. 1043 – 1048.
[38] M. E. Crovella and L. Lipsky, Long-lasting transient conditions in simulations with heavy-tailed workloads, in Proc. 1997 Winter Simulation Conference, December 7-10, vol. Atlanta, GA, USA, Edmonton, Canada, 1997.
[39] A. Feldmann, A. C. Gilbert, P. Huang and W. Willinger, Dynamics of IP traffic: a study of the role of variability and the impact of control, in Proc. Applications, technologies, architectures, and protocols for computer communication, August 30-September 03, Cambridge, Massachusetts, USA, 1999, pp. 301–313.
[40] C. Nuzman, I. Saniee, W. Sweldens and A. Weiss, A compound model for TCP connection arrivals for LAN and WAN applications, Computer Networks: The International Journal of Computer and Telecommunications Networking, vol. 40, no. 3, pp. 319–337, 2002.
[41] B. Ryu and S. Lowen. Fractal Traffic Model for Internet Simulation. In Proc. 5th IEEE Symposium on Computers and Communications (ISCC 2000), 2000.

[42] M. Fras, Methods for the statistical modeling of measured network traffic for simulation purposes, Ph.D. thesis, 2009, Maribor, Slovenia.
[43] http://en.wikipedia.org/wiki/Pareto_distribution.
[44] http://en.wikipedia.org/wiki/Weibull_distribution.
[45] M. Fras, J. Mohorko and Ž. Čučej, Modeling of captured network traffic by the mimic defragmentation process, Simulation: Transactions of The Society for Modeling and Simulation International, San Diego, USA, Published online 20 September 2010.
[46] M. Fras, J. Mohorko and Ž. Čučej, Modeling of measured self-similar network traffic in OPNET simulation tool, Inf. MIDEM, 40(3): 224-231, September 2010.

Part 6

Routing

16

On the Fluid Queue Driven by an Ergodic Birth and Death Process

Fabrice Guillemin[1] and Bruno Sericola[2]
[1]*Orange Labs, Lannion*
[2]*INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex France*

1. Introduction

Fluid models are powerful tools for evaluating the performance of packet telecommunication networks. By masking the complexity of discrete packet based systems, fluid models are in general easier to analyze and yield simple dimensioning formulas. Among fluid queuing systems, those with arrival rates modulated by Markov chains are very efficient to capture the burst structure of packet arrivals, notably in the Internet because of bulk data transfers. By exploiting the Markov property, very efficient numerical algorithms can be designed to estimate performance metrics such as the overflow probability, the delay of a fluid particle or the duration of a busy period.

In the last decade, stochastic fluid models and in particular Markov driven fluid queues, have received a lot of attention in various contexts of system modeling, e.g. manufacturing systems (see Aggarwal et al. (2005)), communication systems (in particular TCP modeling; see vanForeest et al. (2002)) or more recently peer to peer file sharing process (see Kumar et al. (2007)) and economic systems (risk analysis; see Badescu et al. (2005)). Many techniques exist to analyze such systems.

The first studies of such queuing systems can be dated back to the works by Kosten (1984) and Anick et al. (1982), who analyzed fluid models in connection with statistical multiplexing of several identical exponential on-off input sources in a buffer. The above studies mainly focused on the analysis of the stationary regime and have given rise to a series of theoretical developments. For instance, Mitra (1987) and Mitra (1988) generalize this model by considering multiple types of exponential on-off inputs and outputs. Stern & Elwalid (1991) consider such models for separable Markov modulated rate processes which lead to a solution of the equilibrium equations expressed as a sum of terms in Kronecker product form. Igelnik et al. (1995) derive a new approach, based on the use of interpolating polynomials, for the computation of the buffer overflow probability.

Using the Wiener-Hopf factorization of finite Markov chains, Rogers (1994) shows that the distribution of the buffer level has a matrix exponential form, and Rogers & Shi (1994) explore algorithmic issues of that factorization. Ramaswami (1999) and da Silva Soares & Latouche (2002), Ahn & Ramaswami (2003) and da Silva Soares & Latouche (2006) respectively exhibit

and exploit the similarity between stationary fluid queues in a finite Markovian environment and quasi birth and death processes.

Following the work by Sericola (1998) and that by Nabli & Sericola (1996), Nabli (2004) obtained an algorithm to compute the stationary distribution of a fluid queue driven by a finite Markov chain. Most of the above cited studies have been carried out for finite modulating Markov chains.

The analysis of a fluid queue driven by infinite state space Markov chains has also been addressed in many research papers. For instance, when the driving process is the M/M/1 queue, Virtamo & Norros (1994) solve the associated infinite differential system by studying the continuous spectrum of a key matrix. Adan & Resing (1996) consider the background process as an alternating renewal process, corresponding to the successive idle and busy periods of the M/M/1 queue. By renewal theory arguments, the fluid level distribution is given in terms of integral of Bessel functions. They also obtain the expression of Virtamo and Norros via an integral representation of Bessel functions. Barbot & Sericola (2002) obtain an analytic expression for the joint stationary distribution of the buffer level and the state of the M/M/1 queue. This expression is obtained by writing down the solution in terms of a matrix exponential and then by using generating functions that are explicitly inverted.

In Sericola & Tuffin (1999), the authors consider a fluid queue driven by a general Markovian queue with the hypothesis that only one state has a negative drift. By using the differential system, the fluid level distribution is obtained in terms of a series, which coefficients are computed by means of recurrence relations. This study is extended to the finite buffer case in Sericola (2001). More recently, Guillemin & Sericola (2007) considered a more general case of infinite state space Markov process that drives the fluid queue under some general uniformization hypothesis.

The Markov chain describing the number of customers in the M/M/1 queue is a specific birth and death process. Queueing systems with more general modulating infinite Markov chain have been studied by several authors. For instance, van Dorn & Scheinhardt (1997) studied a fluid queue fed by an infinite general birth and death process using spectral theory.

Besides the study of the stationary regime of fluid queues driven by finite or infinite Markov chains, the transient analysis of such queues has been studied by using Laplace transforms by Kobayashi & Ren (1992) and Ren & Kobayashi (1995) for exponential on-off sources. These studies have been extended to the Markov modulated input rate model by Tanaka et al. (1995). Sericola (1998) has obtained a transient solution based on simple recurrence relations, which are particularly interesting for their numerical properties. More recently, Ahn & Ramaswami (2004) use an approach based on an approximation of the fluid model by the amounts of work in a sequence of Markov modulated queues of the quasi birth and death type. When the driving Markov chain has an infinite state space, the transient analysis is more complicated. Sericola et al. (2005) consider the case of the M/M/1 queue by using recurrence relations and Laplace transforms.

In this paper, we analyze the transient behavior of a fluid queue driven by a general ergodic birth and death process using spectral theory in the Laplace transform domain. These results are applied to the stationary regime and to the busy period analysis of that fluid queue.

2. Model description

2.1 Notation and fundamental system

Throughout this paper, we consider a queue fed by a fluid traffic source, whose instantaneous transmitting bit rate is modulated by a general birth and death process (Λ_t) taking values in $\mathbb{N} = \{0, 1, 2, \ldots\}$. The input rate is precisely $r(\Lambda_t)$, where r is a given increasing function from $\mathbb{N}$ into $\mathbb{R}$.

The birth and death process (Λ_t) is characterized by the infinitesimal generator given by the infinite matrix

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & . & . \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & . & . \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & . \\ . & . & . & . & . \end{pmatrix}, \tag{1}$$

where $\lambda_i > 0$ for $i \geq 0$ is the transition rate from state i to state $i+1$ and $\mu_j > 0$ for $j \geq 1$ is the transition rate from state j to state $j-1$.

We assume that the birth and death process (Λ_t) is ergodic, which amounts to assuming (see Asmussen (1987) for instance) that

$$\sum_{i=0}^{\infty} \frac{1}{\lambda_i \pi_i} = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \pi_i < \infty, \tag{2}$$

where the quantities π_i are defined by:

$$\pi_0 = 1 \quad \text{and} \quad \pi_i = \frac{\lambda_0 \ldots \lambda_{i-1}}{\mu_1 \ldots \mu_i}, \quad \text{for} i \geq 1.$$

Under the above assumption, the birth and death process (Λ_t) has a unique invariant probability measure: in steady state, the probability of being in state i is

$$p(i) = \frac{\pi_i}{\sum_{j=0}^{\infty} \pi_j}.$$

Let $p_0(i)$ denote, for $i \geq 0$, the probability that the birth and death process (Λ_t) is in state i at time 0, i.e., $\mathbb{P}(\Lambda_0 = i) = p_0(i)$. Note that if $p_0(i) = p(i)$ for all $i \geq 0$, then $\mathbb{P}(\Lambda_t = i) = p(i)$ for all $t \geq 0$ and $i \geq 0$.

We assume that the queue under consideration is drained at constant rate $c > 0$. Furthermore, we assume that $r(i) > c$ when i is greater than a fixed $i_0 > 0$ and that $r(i) < c$ for $0 \leq i \leq i_0$. (It is worth noting that we assume that $r(i) \neq c$ for all $i \geq 0$ in order to exclude states with no drift and thus to avoid cumbersome special cases.) In addition, the parameters c and $r(i)$ are such that

$$\rho = \sum_{i=0}^{\infty} \frac{r(i)}{c} p(i) < 1 \tag{3}$$

so that the system is stable. The quantity $r_i = r(i) - c$ is either positive or negative and is the net input rate when the modulating process (Λ_t) is in state i.

Let X_t denote the buffer content at time t. The process (X_t) satisfies the following evolution equation: for $t \geq 0$,

$$\frac{dX_t}{dt} = \begin{cases} r(\Lambda_t) - c & \text{if } X_t > 0 \text{ or } r(\Lambda_t) > c, \\ 0 & \text{if } X_t = 0 \text{ and } r(\Lambda_t) \leq c. \end{cases} \tag{4}$$

Let $f_i(t,x)$ denote the joint probability density function defined by

$$f_i(t,x) = \frac{\partial}{\partial x}\mathbb{P}(\Lambda_t = i, X_t \leq x).$$

As shown in Sericola (1998), on top of its usual jump at point $x = 0$, when $X_0 = x_0 \geq 0$, the distribution function $\mathbb{P}(\Lambda_t = i, X_t \leq x)$ has a jump at points $x = x_0 + r_i t$, for t such that $x_0 + r_i t > 0$, which corresponds to the case when the Markov chain $\{\Lambda_t\}$ starts and remains during the whole interval $[0,t)$ in state i.

We focus in the rest of the paper on the probability density function $f_i(t,x)$ for $x > 0$ along with its usual jump at point $x = 0$. A direct consequence of the evolution equation (4) is the forward Chapman-Kolmogorov equations satisfied by $(f_i(t,x),\ x \geq 0, i \in \mathbb{N})$, which form the fundamental system to be solved.

Proposition 1 (Fundamental system)**.** *The functions $(x,t) \to f_i(t,x)$ for $i \in \mathbb{N}$ satisfy the differential system (in the sense of distributions):*

$$\frac{\partial f_i}{\partial t} = -r_i \frac{\partial}{\partial x}\left(\left(\mathbb{1}_{\{i>i_0\}} + \mathbb{1}_{\{i\leq i_0\}}\mathbb{1}_{\{x>0\}}\right) f_i\right) - (\lambda_i + \mu_i) f_i + \lambda_{i-1} f_{i-1} + \mu_{i+1} f_{i+1}, \tag{5}$$

with the convention $\lambda_{-1} = 0$, $f_{-1} \equiv 0$ and $f_i(t,x) = 0$ for $x < 0$.

Note that the differential system (5) holds for the density probability functions $f_i(t,x)$. The differential system considered in Parthasarathy et al. (2004) and van Dorn & Scheinhardt (1997) governs the probability distribution functions $\mathbb{P}(X_t \leq x, \Lambda_t = i), i \geq 0$. The differential system (5) is actually the equivalent of Takács' integro-differential formula for the $M/G/1$ queue, see Kleinrock (1975). The resolution of this differential system is addressed in the next section.

2.2 Basic matrix Equation

Introduce the double Laplace transform

$$F_i(s,\xi) = \int_{0^-}^{\infty}\int_{0^-}^{\infty} e^{-st-\xi x} f_i(t,x)dtdx = \int_0^{\infty} e^{-st}\mathbb{E}\left(^{-\xi X_t}\mathbb{1}_{\{\Lambda_t = i\}}\right)dt$$

and define the functions $f_i^{(0)}(\xi)$ and $h_i(s)$ for $i \in \mathbb{N}$ as follows

$$f_i^{(0)}(\xi) = \int_0^{\infty} e^{-x\xi}\mathbb{P}\{\Lambda_0 = i, X_0 \in dx\}$$
$$h_i(s) = \int_0^{\infty} e^{-st}\mathbb{P}\{\Lambda_t = i, X_t = 0\}dt.$$

The functions $f_i^{(0)}$ are related to the initial conditions of the system and are known functions. For $i > i_0$, we have $\mathbb{P}\{\Lambda_t = i, X_t = 0\} = 0$, which implies that $h_i(s) = 0$, for $i > i_0$. On the contrary, for $i \leq i_0$, the functions h_i are unknown and have to be determined by taking into account the dynamics of the system.

By taking Laplace transforms in Equation (5), we obtain the following result.

Proposition 2. *Let $F(s,\xi)$, $f^{(0)}$, and $h(s)$ be the infinite column vectors, which components are $F_i(s,\xi)/\pi_i$, $f_i^{(0)}/\pi_i$, and $h_i(s)/\pi_i$ for $i \geq 0$, respectively. Then, these vectors satisfy the matrix equation*

$$(s\mathbb{I} + \xi R - A)F(s,\xi) = f^{(0)}(\xi) + \xi R h(s), \tag{6}$$

where $\mathbb{I}$ is the identity matrix, A is the infinitesimal generator of the birth and death process $\{\Lambda_t\}$ defined by Equation (1)*, and R is the diagonal matrix with diagonal elements r_i, $i \geq 0$.*

Proof. Taking the Laplace transform of $\partial f_i/\partial t$ gives rise to the term $sF_i - f_i^{(0)}$. In the same way, taking the Laplace transform of $\partial(\mathbb{1}_{\{x>0\}}f_i)/\partial x$ yields the term $\xi F_i - \xi h_i$. Hence, taking Laplace transforms in Equation (5) and dividing all terms by π_i gives, for $i \geq 0$,

$$s\frac{F_i}{\pi_i} - \frac{f_i^{(0)}}{\pi_i} = -r_i\xi\frac{F_i}{\pi_i} + r_i\xi\frac{h_i}{\pi_i} - (\lambda_i + \mu_i)\frac{F_i}{\pi_i} + \lambda_i\frac{F_{i+1}}{\pi_{i+1}} + \mu_i\frac{F_{i-1}}{\pi_{i-1}},$$

which can be rewritten in matrix form as Equation (6) □

When we consider the stationary regime of the fluid queue, we have to set $f^{(0)}(\xi) \equiv 0$ and eliminate the term $s\mathbb{I}$ in Equation (6), which then becomes

$$(\xi R - A)F(\xi) = \xi R h, \tag{7}$$

where h is the vector, which ith component is $h_i = \lim_{t\longrightarrow\infty}\mathbb{P}\{\Lambda_t = i, X_t = 0\}/\pi_i$ and $F(\xi)$ is the vector, which ith component is $\mathbb{E}\left[e^{-\xi X_t}\mathbb{1}_{\{\Lambda_t=i\}}\right]/\pi_i$. This is the Laplace transform version of Equation (12) by van Dorn & Scheinhardt (1997), which addresses the resolution of Equation (7).

3. Resolution of the fundamental system

In this section, we show how Equation (6) can be solved. For this purpose, we analyze the structure of this equation and in a first step, we prove that the functions $F_i(s,\xi)$ can be expressed in terms of the function $F_{i_0}(s,\xi)$. (Recall that the index i_0 is the greatest integer such that $r(i) - c < 0$ and that for $i \geq i_0 + 1$, $r(i) > c$.). The proof greatly relies on the spectral properties of some operators defined in adequate Hilbert spaces.

3.1 Basic orthogonal polynomials

In the following, we use the orthogonal polynomials $Q_i(s;x)$ defined by recursion: $Q_0(s;x) \equiv 1$, $Q_1(s;x) = (s + \lambda_0 - r_0 x)/\lambda_0$ and for $i \geq 1$,

$$\frac{\lambda_i}{|r_i|}Q_{i+1}(s;x)+\left(x-\frac{s+\lambda_i+\mu_i}{|r_i|}\right)Q_i(s;x)+\frac{\mu_i}{|r_i|}Q_{i-1}(s;x)=0. \tag{8}$$

By suing Favard's criterion (see Askey (1984) for instance), it is easily checked that the polynomials $Q_i(s;x)$ for $i \geq 0$ form an orthogonal polynomial system.

The polynomials $\frac{\lambda_0 \ldots \lambda_{i-1}}{|r_0 \ldots r_{i-1}|}Q_i(s;-z)$, $i \geq 0$ are the successive denominators of the continued fraction

$$\mathcal{F}^e(s;z)=\cfrac{1}{z+\frac{s+\lambda_0}{|r_0|}-\cfrac{\frac{\mu_1\lambda_0}{|r_0r_1|}}{z+\frac{s+\lambda_1+\mu_1}{|r_1|}-\cfrac{\frac{\mu_2\lambda_1}{|r_2r_1|}}{z+\frac{s+\lambda_2+\mu_2}{|r_2|}-\ddots}}}$$

which is itself the even part of the continued fraction

$$\mathcal{F}(s;z)=\cfrac{\alpha_1(s)}{z+\cfrac{\alpha_2(s)}{1+\cfrac{\alpha_3(s)}{z+\cfrac{\alpha_4(s)}{1+\ddots}}}}, \tag{9}$$

where the coefficients $\alpha_k(s)$ are such that $\alpha_1(s)=1$, $\alpha_2(s)=(s+\lambda_0)/|r_0|$, and for $k \geq 1$,

$$\alpha_{2k}(s)\alpha_{2k+1}(s)=\frac{\lambda_{k-1}\mu_k}{|r_{k-1}r_k|}, \quad \alpha_{2k+1}(s)+\alpha_{2(k+1)}(s)=\frac{s+\lambda_k+\mu_k}{|r_k|}. \tag{10}$$

We have the following property, which is proved in Appendix A.

Lemma 1. *The continued fraction $\mathcal{F}(s;z)$ defined by Equation* (9) *is a converging Stieltjes fraction for all $s \geq 0$.*

As a consequence of the above lemma, there exists a unique bounded, increasing function $\psi(s;x)$ in variable x such that

$$\mathcal{F}(s;z)=\int_0^\infty \frac{1}{z+x}\psi(s;dx).$$

The polynomials $Q_n(s;x)$ are orthogonal with respect to the measure $\psi(s;dx)$ and satisfy the orthogonality relation

$$\int_0^\infty Q_i(s;x)Q_j(s;x)\psi(s;dx)=\frac{|r_0|}{|r_i|\pi_i}\delta_{i,j} \tag{11}$$

As a consequence, it is worth noting that the polynomial $Q_i(s;x)$ has i real, simple and positive roots.

It is possible to associate with the polynomials $Q_i(s,x)$ a new class of orthogonal polynomials, referred to as associated polynomials and denoted by $Q_i(i_0+1;s;x)$ and satisfying the

recurrence relations: $Q_0(i_0+1;s;x) = 1$, $Q_1(i_0+1;s;x) = (s+\lambda_{i_0+1+i}+\mu_{i_0+1+i} - r_{i_0+1+i}x)/\lambda_{i_0+1+i}$ and, for $i \geq 0$,

$$\frac{\lambda_{i_0+1+i}}{r_{i_0+1+i}} Q_{i+1}(i_0+1;s;x) + \left(x - \frac{s+\lambda_{i_0+1+i}+\mu_{i_0+1+i}}{r_{i_0+1+i}}\right) Q_i(i_0+1;s;x) + \frac{\mu_{i_0+1+i}}{r_{i_0+1+i}} Q_{i-1}(i_0+1;s;x) = 0. \quad (12)$$

The polynomials $Q_i(i_0+1;s;z)$ are related to the denominator of the continued fraction

$$\mathcal{F}^e_{i_0}(z) = \cfrac{1}{z + \frac{s+\lambda_{i_0+1}+\mu_{i_0+1}}{r_{i_0+1}} - \cfrac{\frac{\lambda_{i_0+1}\mu_{i_0+2}}{r_{i_0+1}r_{i_0+2}}}{z + \frac{s+\lambda_{i_0+2}+\mu_{i_0+2}}{r_{i_0+2}} - \cfrac{\frac{\lambda_{i_0+2}\mu_{i_0+3}}{r_{i_0+2}r_{i_0+3}}}{z + \frac{s+\lambda_{i_0+3}+\mu_{i_0+3}}{|r_{i_0+3}|} - \ddots}}}$$

which is the even part of the continued fraction $\mathcal{F}_{i_0}(z)$ defined by

$$\mathcal{F}_{i_0}(s;z) = \cfrac{\beta_1(s)}{z + \cfrac{\beta_2(s)}{1 + \cfrac{\beta_3(s)}{z + \cfrac{\beta_4(s)}{1 + \ddots}}}}, \quad (13)$$

where the coefficients $\beta_k(s)$ are such that

$$\beta_1(s) = 1, \quad \beta_2(s) = (s+\lambda_{i_0+1}+\mu_{i_0+1})/|r_{i_0+1}|,$$

and for $k \geq 1$,

$$\begin{aligned} \beta_{2k}(s)\beta_{2k+1}(s) &= \frac{\lambda_{i_0+k}\mu_{i_0+k+1}}{r_{i_0+k}r_{i_0+1+k}}, \\ \beta_{2k+1}(s) + \beta_{2(k+1)}(s) &= \frac{s+\lambda_{i_0+1+k}+\mu_{i_0+1+k}}{r_{i_0+1+k}}. \end{aligned} \quad (14)$$

Since the continued fraction $\mathcal{F}(s;z)$ is a converging Stieltjes fraction, it is quite clear that the continued fraction $\mathcal{F}_{i_0}(s;z)$ defined by Equation (13) is a converging Stieltjes fraction for all $s \geq 0$. There exists hence a unique bounded, increasing function $\psi^{[i_0]}(s;x)$ in variable x such that

$$\mathcal{F}_{i_0}(s;z) = \int_0^\infty \frac{1}{z+x} \psi^{[i_0]}(s;dx).$$

The polynomials $Q_i(i_0+1;s;x)$ are orthogonal with respect to the measure $\psi^{[i_0]}(s;dx)$ and satisfy the orthogonality relation

$$\int_0^\infty Q_i(i_0+1;s;x)Q_j(i_0+1;s;x)\psi^{[i_0]}(s;dx) = \frac{r_{i_0+1}\pi_{i_0+1}}{r_{i_0+1+i}\pi_{i_0+1+i}}\delta_{i,j}.$$

3.2 Resolution of the matrix equation

We show in this section how to solve the matrix Equation (6). In a first step, we solve the i_0+1 first linear equations.

Lemma 2. *The functions $F_i(s,\xi)$, for $i \leq i_0$, are related to function $F_{i_0+1}(s,\xi)$ as follows: for $\xi \neq \zeta_k(s)$, $k=0,\ldots,i_0$,*

$$F_i(s,\xi) = \frac{\pi_i}{r_0}\sum_{j=0}^{i_0}(f_j^{(0)}(\xi)+r_j\xi h_j(s))\int_0^\infty \frac{Q_j(s;x)Q_i(s;x)}{\xi-x}\psi_{[i_0]}(s;dx) + \mu_{i_0+1}\frac{\pi_i}{r_0}F_{i_0+1}(s,\xi)\int_0^\infty \frac{Q_{i_0}(s;x)Q_i(s;x)}{\xi-x}\psi_{[i_0]}(s;dx), \quad (15)$$

where the $\zeta_k(s)$ are the roots of the polynomial $Q_{i_0+1}(s;x)$ defined by Equation (8) *and the measure $\psi_{[i_0]}(s;dx)$ is defined by Equation* (45) *in Appendix A.*

Proof. Let $\mathbb{I}_{[i_0]}$, $A_{[i_0]}$ and $R_{[i_0]}$ denote the matrices obtained from the infinite identity matrix, the infinite matrix A defined by Equation (1) and the infinite diagonal matrix R by deleting the rows and the columns with an index greater than i_0, respectively. Denoting by $F_{[i_0]}$, $h_{[i_0]}$ and $f_{[i_0]}$ the finite column vectors which ith components are F_i/π_i, h_i/π_i and $f_i^{(0)}/\pi_i$, respectively for $i=0,\ldots,i_0$, Equation (6) can be written as

$$(s\mathbb{I}_{[i_0]}+\xi R_{[i_0]}-A_{[i_0]})F_{[i_0]} = f_{[i_0]}+\xi R_{[i_0]}h_{[i_0]}+\frac{\lambda_{i_0}}{\pi_{i_0+1}}F_{i_0+1}e_{i_0},$$

where e_{i_0} is the column vector with all entries equal to 0 except the i_0th one equal to 1.

Since $r(i) < c$ for all $i \leq i_0$, the matrix $R_{[i_0]}$ is invertible and the above equation can be rewritten as

$$\left(\xi\mathbb{I}_{[i_0]}+R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]}-A_{[i_0]})\right)F_{[i_0]} = R_{[i_0]}^{-1}f_{[i_0]}+\xi h_{[i_0]}+\frac{\lambda_{i_0}}{r_{i_0}\pi_{i_0+1}}F_{i_0+1}e_{i_0}.$$

From Lemma 6 proved in Appendix B, we know that the operator associated with the finite matrix $(\xi\mathbb{I}_{[i_0]}+R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]}-A_{[i_0]}))$ is selfadjoint in the Hilbert space $H_{i_0}=\mathbb{C}^{i_0+1}$ equipped with the scalar product

$$(c,d)_{i_0} = \sum_{k=0}^{i_0} c_k\overline{d_k}|r_k|\pi_k.$$

The eigenvalues of the operator $(\xi\mathbb{I}_{[i_0]}+R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]}-A_{[i_0]}))$ are the quantities $\xi-\zeta_k(s)$ for $k=0,\ldots,i_0$, where the $\zeta_k(s)$ are the roots of the polynomial $Q_{i_0+1}(s;x)$ defined by Equation (8). Hence, for $\xi \notin \{\zeta_0(s),\ldots,\zeta_{i_0}(s)\}$, we have

$$F_{[i_0]} = \left(\xi\mathbb{I}_{[i_0]}+R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]}-A_{[i_0]})\right)^{-1}R_{[i_0]}^{-1}f_{[i_0]}+\xi\left(\xi\mathbb{I}_{[i_0]}+R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]}-A_{[i_0]})\right)^{-1}h_{[i_0]} + \frac{\lambda_{i_0}}{r_{i_0}\pi_{i_0+1}}F_{i_0+1}\left(\xi\mathbb{I}_{[i_0]}+R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]}-A_{[i_0]})\right)^{-1}e_{i_0}.$$

By introducing the vectors $Q_{[i_0]}(s,\zeta_k(s))$ for $k=0,\dots,i_0$ defined in Appendix B, the column vector e_i with all entries equal to 0 except the ith one equal to 1 can be written as

$$e_j = \frac{|r_j|\pi_j}{|r_0|}\int_0^\infty Q_j(s,x)Q_{[i_0]}(s,x)\psi_{[i_0]}(s;dx)$$

where the measure $\psi_{[i_0]}(s;dx)$ is defined by Equation (45). Since the vectors $Q_{[i_0]}(s,\zeta_k(s))$ are such that

$$\left(\xi\mathbb{I}_{[i_0]} + R^{-1}_{[i_0]}(s\mathbb{I}_{[i_0]} - A_{[i_0]})\right)^{-1} Q_{[i_0]}(s,\zeta_k(s)) = \frac{1}{\xi-\zeta_k(s)}Q_{[i_0]}(s,\zeta_k(s)),$$

we deduce that

$$\left(\xi\mathbb{I}_{[i_0]} + R^{-1}_{[i_0]}(s\mathbb{I}_{[i_0]} - A_{[i_0]})\right)^{-1} e_j = \frac{|r_j|\pi_j}{|r_0|}\int_0^\infty \frac{Q_j(s,x)}{\xi-x}Q_{[i_0]}(s,x)\psi_{[i_0]}(s;dx)$$

Hence, if $f=\sum_{j=0}^{i_0} f_j e_j$, then

$$\left(\xi\mathbb{I}_{[i_0]} + R^{-1}_{[i_0]}(s\mathbb{I}_{[i_0]} - A_{[i_0]})\right)^{-1} f = \sum_{j=0}^{i_0} f_j \frac{|r_j|\pi_j}{|r_0|}\int_0^\infty \frac{Q_j(s,x)}{\xi-x}Q_{[i_0]}(s,x)\psi_{[i_0]}(s;dx)$$

and the ith component of the above vector is

$$\left(\left(\xi\mathbb{I}_{[i_0]} + R^{-1}_{[i_0]}(s\mathbb{I}_{[i_0]} - A_{[i_0]})\right)^{-1} f\right)_i = \sum_{j=0}^{i_0} f_j \frac{|r_j|\pi_j}{|r_0|}\int_0^\infty \frac{Q_j(s,x)Q_i(s,x)}{\xi-x}\psi_{[i_0]}(s;dx)$$

Applying the above identity to the vectors $R^{-1}_{[i_0]}f_{[i_0]}$, $h_{[i_0]}$ and e_{i_0}, Equation (15) follows. □

We now turn to the analysis of the second part of Equation (6).

Lemma 3. *For $s\geq 0$, the functions $F_i(s,\xi)$ are related to function $F_{i_0}(s,\xi)$ by the relation: for $i\geq 0$,*

$$F_{i_0+i+1}(s,\xi) = \lambda_{i_0}\frac{\pi_{i_0+i+1}}{r_{i_0+1}\pi_{i_0+1}}F_{i_0}(s,\xi)\int_0^\infty \frac{Q_i(i_0+1;s;x)}{\xi+x}\psi^{[i_0]}(s;dx)$$

$$+\frac{\pi_{i_0+i+1}}{r_{i_0+1}\pi_{i_0+1}}\sum_{j=0}^{\infty} f^{(0)}_{i_0+j+1}(\xi)\int_0^\infty \frac{Q_j(i_0+1;s;x)Q_i(i_0+1;s;x)}{x+\xi}\psi^{[i_0]}(s;dx), \quad (16)$$

where the measure $\psi^{[i_0]}(s;dx)$ is the orthogonality measure of the associated polynomials $Q_i(i_0+1;s;x)$, $i\geq 0$.

Proof. Let $\mathbb{I}^{[i_0]}$, $A^{[i_0]}$ and $R^{[i_0]}$ denote the matrices obtained from $\mathbb{I}$, A and R by deleting the first (i_0+1) lines and columns, respectively. The infinite matrix $(R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]}-A^{[i_0]})$ induces in the Hilbert space H^{i_0} defined by

$$H^{i_0} = \left\{ (f_n) \in \mathbb{C}^{\mathbb{N}} : \sum_{n=0}^{\infty} |f_n|^2 r_{i_0+n+1} \pi_{i_0+n+1} < \infty \right\}$$

and equipped with the scalar product

$$(f, g) = \sum_{n=0}^{\infty} f_n \overline{g}_n r_{i_0+n+1} \pi_{i_0+n+1},$$

where $\overline{g}_n$ is the conjugate of the complex number g_n, an operator such that for $f \in H^{i_0}$

$$((R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]})f)_n =$$
$$-\frac{\mu_{i_0+1+n}}{r_{i_0+n+1}} f_{n-1} + \frac{s + \lambda_{i_0+n+1} + \mu_{i_0+1+n}}{r_{i_0+n+1}} f_n - \frac{\lambda_{i_0+n+1}}{r_{i_0+n+1}} f_{n+1}.$$

The above operator is symmetric in H^{i_0}. To show that this operator is selfadjoint, we have to prove that the domains of this operator and its adjoint coincide. In Guillemin (2012), it is shown that given the special form of the operator under consideration, this condition is equivalent to the convergence of the Stieltjes fraction defined by Equation (13) and if this is the case, the spectral measure is the orthogonality measure $\psi^{[i_0]}(s; dx)$. Since the continued fraction $\mathcal{F}_{i_0}(s; z)$ is a converging Stieltjes fraction, the above operator is hence selfadjoint.

Let $Q^{[i_0]}(s; x)$ the column vector which ith entry is $Q_i(i_0 + 1; s; x)$. This vector is in H^{i_0} if and only if $\|Q^{[i_0]}(s; x)\|^2 \stackrel{def}{=} (Q^{[i_0]}(s; x), Q^{[i_0]}(s; x)) < \infty$. If it is the case, then the measure $\psi^{[i_0]}(s; dx)$ has an atom at point x with mass $1/\|Q^{[i_0]}(s; x)\|^2$. Otherwise, the vector $Q^{[i_0]}(s; x)$ is not in H^{i_0} but from the spectral theorem we have

$$H^{i_0} = \int^{\oplus} H_x^{i_0} \psi^{[i_0]}(s; dx)$$

where $H_x^{i_0}$ is the vector space spanned by the vector $Q^{[i_0]}(s; x)$ for x in the support of the measure $\psi^{[i_0]}(s; dx)$. In addition, we have the resolvent identity: For $f, g \in H^{i_0}$ and $\xi \in \mathbb{C}$ such that $-\xi$ is not in the support of the measure $\psi^{[i_0]}(s; dx)$,

$$\left(\left(\xi \mathbb{I}^{[i_0]} + (R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]}) \right)^{-1} f, g \right) = \int_0^{\infty} \frac{(f_x, g)}{\xi + x} \psi^{[i_0]}(s; dx). \tag{17}$$

where f_x is the projection on $H_x^{i_0}$ of the vector f.

For $i \geq 0$, let e_i denote the column vector, which ith entry is equal to 1 and the other entries are equal to 0. Denoting by $F^{[i_0]}$ and $\hat{f}^{[i_0]}$ the column vectors which ith components are $F_{i_0+1+i}/\pi_{i_0+1+i}$ and $f^{(0)}_{i_0+1+i}/\pi_{i_0+1+i}$, respectively, Equation (6) can be written as

$$(s\mathbb{I}^{[i_0]} + \xi R^{[i_0]} - A^{[i_0]})F^{[i_0]} = f^{[i_0]} + \frac{\mu_{i_0+1}}{\pi_{i_0}} F_{i_0} e_0,$$

since $h_i(s) \equiv 0$ for $i > i_0$.

Given that $r_i > 0$ for $i > i_0$, the matrix $R^{[i_0]}$ is invertible and the above equation can be rewritten as

$$\left(\xi \mathbb{I}^{[i_0]} + (R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]})\right) F^{[i_0]} = (R^{[i_0]})^{-1} f^{[i_0]} + \frac{\mu_{i_0+1}}{\pi_{i_0}} F_{i_0} \hat{R}^{-1} e_0,$$

The operator $\left(\xi \mathbb{I}^{[i_0]} + (R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]})\right)$ is invertible for ξ such that $-\xi$ is not in the support of the measure $\psi^{[i_0]}(s, dx)$, and we have

$$F^{[i_0]} = \left(\xi \mathbb{I}^{[i_0]} + (R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]})\right)^{-1} (R^{[i_0]})^{-1} f^{[i_0]} + \frac{\mu_{i_0+1}}{r_{i_0+1}\pi_{i_0}} F_{i_0} \left(\xi \mathbb{I}^{[i_0]} + (R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]})\right)^{-1} e_0.$$

By using the spectral identity (17), we can compute F_i for $i > i_0$ as soon as F_{i_0} is known. Indeed, we have

$$F^{[i_0]} = \sum_{j=0}^{\infty} \frac{F_{i_0+1+j}}{\pi_{i_0+1+j}} e_j,$$

and then, for $i \geq i_0 + 1$, by using the fact that $r_{i_0+1+i} F_{i_0+1+i} = (F^{[i_0]}, e_i)$, we have

$$r_{i_0+1+i} F_{i_0+1+i} = \left(\left(\xi \mathbb{I}^{[i_0]} + (R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]})\right)^{-1} (R^{[i_0]})^{-1} f^{[i_0]}, e_i \right) + \frac{\mu_{i_0+1}}{r_{i_0+1}\pi_{i_0}} F_{i_0} \left(\left(\xi \mathbb{I}^{[i_0]} + (R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]})\right)^{-1} e_0, e_i \right).$$

By using the fact that for $j \geq 0$,

$$(e_j)_x = \frac{r_{i_0+j+1}\pi_{i_0+j+1}}{r_{i_0+1}\pi_{i_0+1}} Q_j(i_0+1; s; x) Q^{[i_0]}(s; x),$$

Equation (16) follows by using the resolvent identity (17). □

From the two above lemmas, it turns out that to determine the functions $F_i(s, \xi)$ it is necessary to compute the function $h_i(s)$ for $i = 0, \ldots, i_0 + 1$. For this purpose, let us introduce the non negative quantities $\eta_\ell(s)$, $\ell = 0, \ldots, i_0$, which are the $(i_0 + 1)$ solution to the equation

$$1 - \frac{\lambda_{i_0}\mu_{i_0+1}\pi_{i_0}}{r_{i_0+1}r_0} \mathcal{F}_{i_0}(s; \xi) \int_0^{\infty} \frac{Q_{i_0}(s; x)^2}{\xi - x} \psi_{[i_0]}(s; dx) = 0. \tag{18}$$

Then, we can state the following result, which gives a means of computing the unknown functions $h_j(s)$ for $j = 0, \ldots, i_0$.

Proposition 3. *The functions $h_j(s)$, $j = 0, \ldots, i_0$, satisfy the linear equations: for $\ell = 0, \ldots, i_0$,*

$$\begin{aligned}\frac{\lambda_{i_0}\mathcal{F}_{i_0}(s;\eta_\ell(s))\eta_\ell(s)}{r_{i_0}} &\left(\left(\eta_k(s)\mathbb{I}_{[i_0]} + R^{-1}_{[i_0]}(s\mathbb{I}_{[i_0]} - A_{[i_0]})\right)^{-1} e_{i_0}, h(s)\right)_{i_0} \\ &= \left(\left(\eta_k(s)\mathbb{I}^{[i_0]} + (R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]} - A^{[i_0]})\right)^{-1} e_0, (R^{[i_0]})^{-1} f^{[i_0]}(\eta_k(s))\right) \\ &- \frac{\lambda_{i_0}\mathcal{F}_{i_0}(s;\eta_\ell(s))}{r_{i_0}}\left(\left(\eta_k(s)\mathbb{I}_{[i_0]} + R^{-1}_{[i_0]}(s\mathbb{I}_{[i_0]} - A_{[i_0]})\right)^{-1} e_{i_0}, R^{-1}_{[i_0]} f_{[i_0]}(\eta_k(s))\right)_{i_0}, \quad (19)\end{aligned}$$

where $\mathcal{F}_{i_0}(s;z)$ is the continued fraction (13) *and $f^{[i_0]}((\xi)$ and $f_{[i_0]}(\xi)$ are the vectors, which ith components are equal to $f^{(0)}_{i_0+i+1}(\xi)/\pi_{i_0+i+1}$ and $f^{(0)}_i(\xi)/\pi_i$, respectively.*

Proof. From Equation (16) for $i = i_0 + 1$ and Equation (15) for $i = i_0$, we deduce that

$$\begin{aligned}\left(1 - \frac{\lambda_{i_0}\mu_{i_0+1}\pi_{i_0}}{r_0 r_{i_0+1}}\mathcal{F}_{i_0}(s;\xi)\int_0^\infty \frac{Q_{i_0}(s;x)^2}{\xi - x}\psi_{[i_0]}(s;dx)\right) F_{i_0+1}(s,\xi) = \\ \frac{\lambda_{i_0}\pi_{i_0}}{r_0 r_{i_0+1}}\mathcal{F}_{i_0}(s;\xi)\sum_{j=0}^{i_0}(f^{(0)}_j(\xi) + r_j\xi h_j(s))\int_0^\infty \frac{Q_j(s;x)Q_{i_0}(s;x)}{\xi - x}\psi_{[i_0]}(s;dx) \\ + \frac{1}{r_{i_0+1}}\sum_{j=0}^{\infty} f^{(0)}_{i_0+j+1}(\xi)\int_0^\infty \frac{Q_j(i_0+1;s;x)}{x+\xi}\psi^{[i_0]}(s;dx). \quad (20)\end{aligned}$$

From equation (15), since the Laplace transform $F_i(s,\xi)$ should have no poles for $\xi \geq 0$, the roots $\zeta_k(s)$ for $k = 0, \ldots, i_0$ should be removable singularities and hence for all $i, j, k = 0, \ldots, i_0$

$$\begin{aligned}Q_i(s;\zeta_k(s))\left(\left(f^{(0)}_j(\zeta_k(s)) + r_j\zeta_k(s)h_j(\zeta_k(s))\right)Q_j(s;\zeta_k(s))\right. \\ \left. +\mu_{i_0+1}F_{i_0+1}(s,\zeta_k(s))Q_{i_0}(s,\zeta_k(s))\right) = 0.\end{aligned}$$

By using the interleaving property of the roots of successive orthogonal polynomials, we have $Q_i(s;\zeta_k(s)) \neq 0$ for all $i, k = 0, \ldots, i_0$. Hence, the term between parentheses in the above equation is null and we deduce that the points $\zeta_k(s)$, $k = 0, \ldots, i_0$, are removable singularities in expression (20). The quantities $h_j(s)$, $j = 0, \ldots, i_0$, are then determined by using the fact that the r.h.s. of equation (20) must cancel at points $\eta_k(s)$ for $k = 0, \ldots, i_0$. This entails that for $k = 0, \ldots, i_0$, the terms

$$\begin{aligned}\sum_{j=0}^{\infty} f^{(0)}_{i_0+j+1}(\eta_k(s))\int_0^\infty \frac{Q_j(i_0+1;s;x)}{x+\eta_k(s)}\psi^{[i_0]}(s;dx) \\ + \frac{\lambda_{i_0}\pi_{i_0}\mathcal{F}_{i_0}(s;\eta_k(s))}{r_0}\sum_{j=0}^{i_0} v_j(s)\int_0^\infty \frac{Q_j(s;x)Q_{i_0}(s;x)}{\eta_k(s) - x}\psi_{[i_0]}(s;dx) \quad (21)\end{aligned}$$

must cancel, where

$$v_j(s) = f^{(0)}_j(\eta_k(s)) + \eta_k(s) r_j h_j(s).$$

By using the fact that

$$\int_0^\infty \frac{Q_j(s;x)Q_{i_0}(s;x)}{\eta_k(s)-x}\psi_{[i_0]}(s;dx) = \frac{|r_0|}{|r_{i_0}|\pi_{i_0}|r_j|\pi_j}\left(\left(\eta_k(s)\mathbb{I}_{[i_0]}+R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]}-A_{[i_0]})\right)^{-1}e_{i_0},e_j\right)_{i_0}$$

and

$$\int_0^\infty \frac{Q_j(i_0+1;s;x)}{x+\eta_k(s)}\psi^{[i_0]}(s;dx) = \frac{1}{r_{i_0+1+j}\pi_{i_0+j+1}}\left(\left(\eta_k(s)\mathbb{I}^{[i_0]}+(R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]}-A^{[i_0]})\right)^{-1}e_0,e_j\right),$$

Equation (19) follows. □

By solving the system of linear equations (19), we can compute the unknown functions $h_j(s)$ for $j=0,\dots,i_0$. The function $F_{i_0+1}(s,\xi)$ is then given by

$$\left(1-\frac{\lambda_{i_0}\mu_{i_0+1}\pi_{i_0}}{r_{i_0+1}r_0}\mathcal{F}_{i_0}(s;\xi)\int_0^\infty\frac{Q_{i_0}(s;x)^2}{\xi-x}\psi_{[i_0]}(s;dx)\right)F_{i_0+1}(s,\xi) =$$

$$= \frac{1}{r_{i_0+1}}\left(\left(\xi\mathbb{I}^{[i_0]}+(R^{[i_0]})^{-1}(s\mathbb{I}^{[i_0]}-A^{[i_0]})\right)^{-1}e_0,(R^{[i_0]})^{-1}f^{[i_0]}(\xi)\right)$$

$$-\frac{\lambda_{i_0}\mathcal{F}_{i_0}(s;\xi)}{r_{i_0}r_{i_0+1}}\left(\left(\xi\mathbb{I}_{[i_0]}+R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]}-A_{[i_0]})\right)^{-1}e_{i_0},R_{[i_0]}^{-1}f_{[i_0]}(\xi)+\xi h(s)\right)_{i_0}, \quad (22)$$

The function $F_{i_0}(s,\xi)$ is computed by using equation (22) and equation (15) for $i=i_0$. The other functions $F_i(s,\xi)$ are computed by using Lemmas 2 and 3.

The above procedure can be applied for any value i_0 but expressions are much simpler when $i_0=0$, i.e., when there is only one state with negative net input rate. In that case, we have the following result, when the buffer is initially empty and the birth and death process is in state 1.

Proposition 4. *Assume that $r_0<0$ and $r_i>0$ for $i>0$. When the buffer is initially empty and the birth and death process is in the state 1 at time 0 (i.e., $p_0(i)=\delta_{1,i}$ for all $i\geq 0$), the Laplace transform $h_0(s)$ is given by*

$$h_0(s)=\frac{r_0\eta_0(s)+s+\lambda_0}{\lambda_0\eta_0(s)|r_0|}=\frac{\mu_1\mathcal{F}_0(s;\eta_0(s))}{r_1|r_0|\eta_0(s)}. \quad (23)$$

where $\eta_0(s)$ is the unique positive solution to the equation

$$1-\frac{\lambda_0\mu_1\mathcal{F}_0(s;\xi)}{r_1(s+\lambda_0+r_0\xi)}=0.$$

In addition,

$$F_1(s,\xi)=\frac{\frac{1}{r_1}\left(1+\frac{\lambda_0\xi r_0 h_0(s)}{s+\lambda_0+r_0\xi}\right)\mathcal{F}_0(s;\xi)}{1-\frac{\lambda_0\mu_1}{r_1(s+\lambda_0+r_0\xi)}\mathcal{F}_0(s;\xi)}. \tag{24}$$

Proof. In the case $i_0=0$, the unique root to the equation $Q_1(s;x)$ is $\zeta_0(s)=(s+\lambda_0)|r_0|$. The measure $\psi_{[0]}(s;dx)$ is given by

$$\psi_{[0]}(s;dx)=\delta_{\zeta_0(s)}(dx)$$

and Equation (18) reads

$$1-\frac{\lambda_0\mu_1}{r_1}\mathcal{F}_0(s;\xi)\frac{1}{s+\lambda_0+r_0\xi}=0$$

which has a unique solution $\eta_0(s)>0$. When the buffer is initially empty and the birth and death process is in the state 1 at time 0, we have $f_i^{(0)}(\xi)=\delta_{1,j}$. Then,

$$\begin{aligned}&\left(\left(\eta_0(s)\mathbb{I}^{[0]}+(R^{[0]})^{-1}(s\mathbb{I}^{[0]}-A^{[0]})\right)^{-1}e_0,(R^{[0]})^{-1}f^{[0]}(\eta_0(s))\right)\\&\quad=\frac{1}{r_1\pi_1}\left(\left(\eta_0(s)\mathbb{I}^{[0]}+(R^{[0]})^{-1}(s\mathbb{I}^{[0]}-A^{[0]})\right)^{-1}e_0,e_0\right)=\int_0^\infty\frac{1}{\eta_0(s)+x}\psi^{[0]}(s;dx)\\&\qquad=\mathcal{F}_0(s;\eta_0(s)),\end{aligned}$$

where we have used the resolvent identity (17) and the fact that $(e_0)_x=Q^{[0]}(s;x)$. Moreover,

$$\begin{aligned}&\left(\left(\eta_0(s)\mathbb{I}_{[0]}+R_{[0]}^{-1}(s\mathbb{I}_{[0]}-A_{[0]})\right)^{-1}e_0,R_{[0]}^{-1}f_{[0]}(\eta_0(s))+h(s)\right)_0\\&\qquad=\frac{h_0(s)}{\eta_0(s)+\frac{s+\lambda_0}{r_0}}(e_0,e_0)_0=\frac{h_0(s)|r_0|}{\eta_0(s)+\frac{s+\lambda_0}{r_0}}.\end{aligned}$$

By using Equation (19) for $i_0=0$, Equation (23) follows. Finally, Equation (24) is obtained by using Equation (22). □

4. Analysis of the stationary regime

In this section, we analyze the stationary regime. In this case, we have to take $s=0$ and $f^{(0)}\equiv 0$. To alleviate the notation, we set $\psi_{[i_0]}(0;dx)=\psi_{[i_0]}(dx)$, $\psi^{[i_0]}(0;dx)=\psi^{[i_0]}(dx)$ and $Q_j(0;x)=Q_j(x)$ and $Q_j(i_0+1;0;x)=Q_j(i_0+1;x)$. Equation (20) then reads

$$\begin{aligned}&\left(1-\frac{\lambda_{i_0}\mu_{i_0+1}\pi_{i_0}}{r_{i_0+1}r_0}\mathcal{F}_{i_0}(\xi)\int_0^\infty\frac{Q_{i_0}(x)^2}{\xi-x}\psi_{[i_0]}(dx)\right)F_{i_0+1}(\xi)\\&\qquad=\frac{\lambda_{i_0}\pi_{i_0}\xi\mathcal{F}_{i_0}(\xi)}{r_0r_{i_0+1}}\sum_{j=0}^{i_0}r_jh_j\int_0^\infty\frac{Q_j(x)Q_{i_0}(x)}{\xi-x}\psi_{[i_0]}(dx),\end{aligned} \tag{25}$$

where $h_j = \lim_{t\longrightarrow\infty} \mathbb{P}(\Lambda_t = j, X_t = 0)$, $\mathcal{F}_{i_0}(\xi) = \mathcal{F}_{i_0}(0;\xi)$ and $\mathcal{F}_{i_0+1}(\xi) = \mathcal{F}_{i_0+1}(0;\xi)$.

The continued fraction $\mathcal{F}_{i_0}(\xi)$ has the following probabilistic interpretation:

$$\mu_{i_0+1}\mathcal{F}_{i_0}(\xi)/r_{i_0+1} = \mathbb{E}\left(e^{-\xi\theta_{i_0}}\right)$$

where θ_{i_0} is the passage time of the birth and death process with birth rates $\lambda_n/|r_n|$ and death rates $\mu_n/|r_n|$ from state i_0+1 to state i_0 (see Guillemin & Pinchon (1999) for details). This entails in particular that $\mathcal{F}_{i_0}(0) = r_{i_0+1}/\mu_{i_0+1}$.

Let us first characterize the measure $\psi_{[i_0]}(dx)$. For this purpose, let us introduce the polynomials of the second kind associated with the polynomials $Q_i(x)$. The polynomials of the second kind $P_i(x)$ satisfy the same recursion as the polynomials $Q_i(x)$ but wit the initial conditions $P_0(x) = 0$ and $P_1(x) = |r_0|/\lambda_0$. The even numerators of the continued fraction $\mathcal{F}(z) \stackrel{def}{=} \mathcal{F}(0;z)$, where $\mathcal{F}(s;z)$ is defined by Equation (9), are equal to $\frac{\lambda_0\ldots\lambda_{n-1}}{|r_0\ldots r_{n-1}|}P_n(-z)$ and the even denominators to $\frac{\lambda_0\ldots\lambda_{n-1}}{|r_0\ldots r_{n-1}|}Q_n(-z)$.

Lemma 4. *The spectral measure $\psi_{[i_0]}(dx)$ of the non negative selfadjoint operator $R^{-1}_{[i_0]}A_{[i_0]}$ in the Hilbert space H_{i_0} is such that*

$$\int_0^\infty \frac{1}{z-x}\psi_{[i_0]}(dx) = -\frac{P_{i_0+1}(z)}{Q_{i_0+1}(z)}. \tag{26}$$

The measure $\psi_{[i_0]}(dx)$ is purely discrete with atoms located at the zeros ζ_k, $k = 0,\ldots,i_0$, of the polynomial $Q_{i_0+1}(z)$.

Proof. Let $P_{[i_0]}(z)$ (resp. $Q_{[i_0]}(z)$) denote the column vector, which ith component for $0 \le i \le i_0$ is $P_i(z)$ (resp. $Q_i(z)$). For any $x, z \in \mathbb{C}$, we have

$$\left(z\mathbb{I}_{[i_0]} - R^{-1}_{[i_0]}A_{[i_0]}\right)\left(P_{[i_0]}(z) + xQ_{[i_0]}(z)\right) = e_0 - \frac{\lambda_{i_0}}{|r_{i_0+1}|}\left(P_{i_0+1}(z) + xQ_{i_0+1}(z)\right)e_{i_0}.$$

Hence, if $z \neq \zeta_i$ for $0 \le i \le i_0$, where ζ_i is the ith zero of the polynomial $Q_{i_0+1}(x)$, and if we take $x = -P_{i_0+1}(z)/Q_{i_0+1}(z)$, we see that

$$\left(z\mathbb{I}_{[i_0]} - R^{-1}_{[i_0]}A_{[i_0]}\right)^{-1} e_0 = P_{[i_0]}(z) - \frac{P_{i_0+1}(z)}{Q_{i_0+1}(z)}Q_{[i_0]}(z).$$

From the spectral identity for the operator $R^{-1}_{[i_0]}A_{[i_0]}$ (similar to Equation (17)), we have

$$\left(\left(z\mathbb{I}_{[i_0]} - R^{-1}_{[i_0]}A_{[i_0]}\right)^{-1} e_0, e_0\right)_{i_0} = \int_0^\infty \frac{((e_0)_x, e_0)_{i_0}}{z-x}\psi_{[i_0]}(dx) = -\frac{P_{i_0+1}(z)}{Q_{i_0+1}(z)}|r_0|.$$

Since $(e_0)_x = Q_{[i_0]}(x)$ because of the orthogonality relation (11), Equation (26) immediately follows. □

By using the above lemma, we can show that the smallest solution to the equation

$$1 - \frac{\lambda_{i_0}\mu_{i_0+1}\pi_{i_0}}{r_{i_0+1}r_0}\mathcal{F}_{i_0}(\xi)\int_0^\infty \frac{Q_{i_0}(x)^2}{\xi - x}\psi_{[i_0]}(dx) = 0 \tag{27}$$

is $\eta_0 = 0$. The above equation is the stationary version of Equation (18).

Lemma 5. *The solutions η_j, $j = 0, \ldots, i_0$, to Equation* (27) *are such that $\eta_0 = 0 < \eta_1 < \ldots < \eta_{i_0}$. For $\ell = 1, \ldots, i_0$, η_ℓ is solution to equation*

$$1 = \frac{\mu_{i_0+1}}{r_{i_0+1}}\mathcal{F}_{i_0}(\xi)\frac{Q_{i_0}(\xi)}{Q_{i_0+1}(\xi)}. \tag{28}$$

Proof. The fraction $P_{i_0+1}(z)/Q_{i_0+1}(z)$ is a terminating fraction and from Equation (26), we have

$$\frac{P_{i_0+1}(-z)}{Q_{i_0+1}(-z)} = \int_0^\infty \frac{1}{z+x}\psi_{[i_0]}(dx).$$

On the one hand, by applying Theorem 12.11d of Henrici (1977) to this fraction, we have

$$\frac{P_{i_0+1}(-z)}{Q_{i_0+1}(-z)} - \frac{P_{i_0}(-z)}{Q_{i_0}(-z)} = \int_0^\infty \frac{Q_{i_0}(x)^2}{Q_{i_0}(-z)^2}\frac{\psi_{[i_0]}(dx)}{z+x}. \tag{29}$$

On the other hand, by using the fact that

$$\frac{P_{i_0+1}(-z)}{Q_{i_0+1}(-z)} - \frac{P_{i_0}(-z)}{Q_{i_0}(-z)} = \frac{|r_0|}{\lambda_{i_0}\pi_{i_0}Q_{i_0+1}(-z)Q_{i_0}(-z)}, \tag{30}$$

we deduce that

$$\int_0^\infty \frac{Q_{i_0}(x)^2}{x}\psi_{[i_0]}(dx) = \frac{|r_0|}{\lambda_{i_0}\pi_{i_0}},$$

since $Q_i(0) = 1$ for all $i \geq 0$. In addition, by using the fact that $\mathcal{F}_{i_0}(0) = r_{i_0+1}/\mu_{i_0+1}$, we deduce that the smallest root of Equation (27) is $\eta_0 = 0$. The other roots are positive. Equation (27) can be rewritten as Equation (28) by using Equations (29) and (30). □

Note that by using the same arguments as above, we can simplify Equation (18). As a matter of fact, we have

$$\frac{P_{i_0+1}(s,-z)}{Q_{i_0+1}(s,-z)} - \frac{P_{i_0}(s,-z)}{Q_{i_0}(s,-z)} = \frac{|r_0|}{\lambda_{i_0}\pi_{i_0}Q_{i_0+1}(s,-z)Q_{i_0}(s,-z)},$$

so Equation (18) becomes

$$1 = \frac{\mu_{i_0+1}}{r_{i_0+1}}\mathcal{F}_{i_0}(s,\xi)\frac{Q_{i_0}(s,\xi)}{Q_{i_0+1}(s,\xi)}. \tag{31}$$

The quantities h_i are evaluated by using the normalizing condition $\sum_{i=0}^{i_0} h_i = 1 - \rho$, where ρ is defined by Equation (3), and by solving the i_0 linear equations

$$\ell = 1, \ldots, i_0, \quad \left((\eta_\ell \mathbb{I} - R_{[i_0]}^{-1}A_{[i_0]})^{-1}e_{i_0}, h\right)_{i_0} = 0, \tag{32}$$

where h is the vector which ith component is h_i/π_i. Once the quantities h_i, $i = 0, \ldots, i_0$ are known, the function $F_{i_0+1}(\xi)$ is computed by using relation (25). The function $F_{i_0}(\xi)$ is computed by using the relation

$$F_{i_0+1}(\xi) = \frac{\lambda_{i_0}}{r_{i_0+1}} F_{i_0}(\xi)\mathcal{F}_{i_0}(\xi).$$

This allows us to determine the functions $F_{i_0+1}(\xi)$ and $F_{i_0}(\xi)$. The functions $F_i(\xi)$ for $i = 0, \ldots, i_0$ are computed by using Equation (15) for $s = 0$ and $f^{(0)} \equiv 0$. The functions $F_i(\xi)$ for $i > i_0$ are computed by using Equation (16) for $s = 0$ and $f^{(0)} \equiv 0$. This leads to the following result.

Proposition 5. *The Laplace transform of the buffer content X in the stationary regime is given by*

$$\mathbb{E}\left(e^{-\xi X}\right) = \sum_{i=0}^{\infty} F_i(\xi) = \frac{1}{r_0}\sum_{j=0}^{i_0} r_j \xi h_j \int_0^{\infty} \frac{Q_j(x)\Pi(x)}{\xi - x}\psi_{[i_0]}(dx)$$
$$+ \frac{\lambda_{i_0}}{r_{i_0+1}} F_{i_0}(\xi)\left(\frac{\mu_{i_0+1}}{r_0}\mathcal{F}_{i_0}(\xi)\int_0^{\infty}\frac{Q_{i_0}(x)\Pi(x)}{\xi - x}\psi_{[i_0]}(dx) + \frac{1}{\pi_{i_0+1}}\int_0^{\infty}\frac{\Pi_{i_0}(x)}{x+\xi}\psi^{[i_0]}(dx)\right) \quad (33)$$

with

$$\Pi(x) = \sum_{i=0}^{i_0} \pi_i Q_i(x),$$
$$\Pi_{i_0}(x) = \sum_{i=0}^{\infty} \pi_{i_0+1+i} Q_i(i_0+1; x),$$
$$F_{i_0}(\xi) = \frac{\dfrac{\pi_{i_0}}{r_0}\displaystyle\sum_{j=0}^{i_0} r_j \xi h_j \int_0^{\infty}\frac{Q_j(x)Q_{i_0}(x)}{\xi - x}\psi_{[i_0]}(dx)}{1 - \dfrac{\lambda_{i_0}\mu_{i_0+1}\pi_{i_0}}{r_0 r_{i_0+1}}\mathcal{F}_{i_0}(\xi)\displaystyle\int_0^{\infty}\frac{Q_{i_0}(x)^2}{\xi - x}\psi_{[i_0]}(dx)}.$$

In the case when there is only one state with negative drift, the above result can be simplified as follows.

Corollary 1. *When there is only one state with negative drift, the Laplace transform of the buffer content is given by*

$$\mathbb{E}\left(e^{-\xi X}\right) = \frac{\xi(1-\rho)r_0}{r_0\xi + \lambda_0 - \frac{\lambda_0\mu_1}{r_1}\mathcal{F}_0(\xi)}\left(1 + \frac{\lambda_1}{r_1}\int_0^{\infty}\frac{\Pi_0(x)}{x+\xi}\psi^{[0]}(dx)\right). \quad (34)$$

Proof. Since $\psi_{[0]}(dx) = \delta_{\zeta_0}(dx)$ with $\zeta_0 = \lambda_0/|r_0|$ and $\Pi(x) = 1$, we have

$$\int_0^{\infty}\frac{\Pi(x)}{\xi - x}\psi_{[i_0]}(dx) = \frac{r_0}{r_0\xi + \lambda_0}.$$

Moreover, we have $h_0 = 1 - \rho$ and then

$$F_0(\xi) = \frac{(1-\rho)\xi r_0}{r_0\xi + \lambda_0 - \frac{\lambda_0\mu_1}{r_1}\mathcal{F}_0(\xi)}.$$

Simple algebra then yields equation (34). □

By examining the singularities in Equation (34), it is possible to determine the tail of the probability distribution of the buffer content in the stationary regime. The asymptotic behavior greatly depends on the properties of the polynomials $Q_i(x)$ and their associated spectral measure.

5. Busy period

In this section, we are interested in the duration of a busy period of the fluid reservoir. At the beginning of a busy period, the buffer is empty and the modulating process is in state $i_0 + 1$. More generally, let us introduce the occupation duration B which is the duration the server is busy up to an idle period. The random variable B depends on the initial conditions and we define the conditional probability distribution

$$H_i(t,x) = \mathbb{P}(B \leq t \mid \Lambda_0 = i, X_0 = x).$$

The probability distribution function of a busy period β of the buffer is clearly given by

$$\mathbb{P}(\beta \leq t) = H_{i_0+1}(t,0). \tag{35}$$

It is known in Barbot et al. (2001) that for $t > 0$ and $x > 0$, $H_i(t,x)$ satisfies the following partial differential equations

$$\frac{\partial}{\partial t}H_i(t,x) - r_i\frac{\partial}{\partial x}H_i(t,x) = -\mu_i H_{i-1}(t,x) + (\lambda_i + \mu_i)H_i(t,x) - \lambda_i H_{i+1}(t,x) \tag{36}$$

with the boundary conditions

$$\begin{aligned} H_i(t,0) &= 1 \quad \text{if} \quad t \geq 0,\ r_i \leq 0, \\ H_i(0,x) &= 0 \quad \text{if} \quad x > 0, \\ H_i(0,0) &= 0 \quad \text{if} \quad r_i > 0. \end{aligned}$$

Define then conditional Laplace transform

$$\theta_i(u,x) = \mathbb{E}\left(e^{-uB} \mid \Lambda_0 = i, Q_0 = x\right).$$

By taking Laplace transforms in Equation (36), we have

$$r_i\frac{\partial}{\partial x}\theta_i(u,x) = u\theta_i(u,x) - \mu_i\theta_{i-1}(u,x) + (\lambda_i + \mu_i)\theta_i(u,x) - \lambda_i\theta_{i+1}(u,x)$$

By introducing the conditional double Laplace transform

$$\tilde{\theta}_i(u,\xi) = \int_0^\infty e^{-\xi x}\theta_i(u,x)dx.$$

we obtain for $i \geq 0$

$$r_i\xi\tilde{\theta}_i(u,\xi) - r_i\theta_i(u,0) = u\tilde{\theta}_i(u,\xi) - \mu_i\tilde{\theta}_{i-1}(u,\xi) + (\lambda_i+\mu_i)\tilde{\theta}_i(u,\xi) - \lambda_i\tilde{\theta}_{i+1}(u,\xi)$$

By introducing the infinite vector $\Theta(u,\xi)$, which ith component is $\tilde{\theta}_i(u,\xi)$, the above equations can be rewritten in matrix form as

$$\xi R\Theta(u,\xi) = RT(u) + (u\mathbb{I} - A)\Theta(u,\xi), \tag{37}$$

where $T(u)$ is the vector which ith component is equal to $\theta_i(u,0)$. We clearly have $\theta_i(u,0) = 1$ for $i = 0,\ldots,i_0$. For the moment, the functions $\theta_i(u,0)$ for $i > i_0$ are unknown functions.

Equation (37) can be solved by using the same technique as in Section 3. In the following, we assume that the measure $\psi^{[i_0]}(s;dx)$ has a discrete spectrum with atoms located at points $\chi_k(s) > 0$ for $k \geq 0$. This assumption is satisfied for instance when the measure $\psi(s;dx)$ has a discrete spectrum (see Guillemin & Pinchon (1999) for details). Under this assumption, let $\chi_k(s) > 0$ for $k \geq 0$ be the solutions to the equation

$$\frac{\mu_{i_0+1}}{r_{i_0+1}}\frac{Q_{i_0}(u;-\xi)}{Q_{i_0+1}(u;-\xi)}\mathcal{F}_{i_0}(u,-\xi) = 1.$$

Proposition 6. *The Laplace transforms $\theta_{i_0+1+j}(u,0)$ for $j \geq 0$ satisfy the following linear equations:*

$$\frac{1}{r_{i_0+1}\pi_{i_0+1}}\frac{Q_{i_0}(u;-\xi)}{Q_{i_0+1}(u;-\xi)}\sum_{j=0}^{\infty} r_{i_0+1+j}\pi_{i_0+1+j}\theta_{i_0+1+j}(u,0)\int_0^\infty \frac{Q_j(i_0+1;u;x)}{\xi - x}\psi^{[i_0]}(u;dx)$$
$$+ \frac{1}{|r_0|}\sum_{j=0}^{i_0}|r_j|\pi_j\int_0^\infty \frac{Q_{i_0}(u;x)Q_j(u;x)}{\xi + x}\psi_{[i_0]}(u;dx) = 0 \tag{38}$$

for $\xi \in \{\chi_k(s), k \geq 0\}$.

Proof. Equation (37) can be split into two parts. The first part reads

$$\left(\xi\mathbb{I}_{[i_0]} - R_{[i_0]}^{-1}\left(u\mathbb{I}_{[i_0]} - A_{[i_0]}\right)\right)\Theta_{[i_0]} = e_{[i_0]} - \frac{\lambda_{i_0}}{r_{i_0}}\tilde{\theta}_{i_0+1}(u,\xi)e_{i_0}, \tag{39}$$

where $e_{[i_0]}$ is the finite vector with all entries equal to 1 for $i = 0,\ldots,i_0$ and $\Theta_{[i_0]}$ is the finite vector, which ith entry is $\tilde{\theta}_i(u,\xi)$ for $i = 0,\ldots,i_0$. The second part of the equation is

$$\left(\xi\mathbb{I}^{[i_0]} - \left(R^{[i_0]}\right)^{-1}\left(u\mathbb{I}^{[i_0]} - A^{[i_0]}\right)\right)\Theta^{[i_0]} = T^{[i_0]} - \frac{\mu_{i_0+1}}{r_{i_0+1}}\tilde{\theta}_{i_0}(u,\xi)e_0, \tag{40}$$

where the vector $T^{[i_0]}$ (resp. $\Theta^{[i_0]}$) has entries equal to $\theta_{i_0+1+i}(u,0)$ (resp. $\tilde{\theta}_{i_0+1+i}(u,\xi)$) for $i \geq 0$.

By adapting the proofs in Section 3, we have for $i = 0, \ldots, i_0$

$$\tilde{\theta}_i(u,\xi) = \frac{1}{|r_0|}\sum_{j=0}^{i_0}|r_j|\pi_j\int_0^\infty \frac{Q_i(u;x)Q_j(u;x)}{\xi+x}\psi_{[i_0]}(u;dx) + \frac{\mu_{i_0+1}\pi_{i_0+1}}{|r_0|}\tilde{\theta}_{i_0+1}(u,\xi)\int_0^\infty \frac{Q_{i_0}(u;x)Q_i(s;x)}{\xi+x}\psi_{[i_0]}(u;dx), \quad (41)$$

and for $i \geq 0$

$$\tilde{\theta}_{i_0+i+1}(u,\xi) = -\frac{\mu_{i_0+1+i}}{r_{i_0+1}}\tilde{\theta}_{i_0}(u,\xi)\int_0^\infty \frac{Q_i(i_0+1;u;x)}{\xi-x}\psi^{[i_0]}(u;dx) + \frac{1}{r_{i_0+1}\pi_{i_0+1}}\sum_{j=0}^{\infty} r_{i_0+1+j}\pi_{i_0+1+j}\theta_{i_0+1+j}(u,0)\int_0^\infty \frac{Q_j(i_0+1;u;x)Q_i(i_0+1;u;x)}{\xi-x}\psi^{[i_0]}(u;dx) \quad (42)$$

By using Equation 41 for $i = i_0$ and Equation (42) for $i = 0$, we obtain

$$\left(1 - \frac{\mu_{i_0+1}}{r_{i_0+1}}\frac{Q_{i_0}(u;-\xi)}{Q_{i_0+1}(u;-\xi)}\mathcal{F}_{i_0}(u,-\xi)\right)\tilde{\theta}_{i_0}(u,\xi) = \frac{1}{|r_0|}\sum_{j=0}^{i_0}|r_j|\pi_j\int_0^\infty \frac{Q_{i_0}(u;x)Q_j(u;x)}{\xi+x}\psi_{[i_0]}(u;dx) + \frac{1}{r_{i_0+1}\pi_{i_0+1}}\frac{Q_{i_0}(u;-\xi)}{Q_{i_0+1}(u;-\xi)}\sum_{j=0}^{\infty} r_{i_0+1+j}\pi_{i_0+1+j}\theta_{i_0+1+j}(u,0)\int_0^\infty \frac{Q_j(i_0+1;u;x)}{\xi-x}\psi^{[i_0]}(u;dx)$$

where we have used the fact

$$\int_0^\infty \frac{Q_{i_0}(u;x)^2}{\xi+x}\psi_{[i_0]}(u;dx) = \frac{|r_0|}{\lambda_{i_0}\pi_{i_0}}\frac{Q_{i_0}(u;-\xi)}{Q_{i_0+1}(u;-\xi)}$$

and

$$\int_0^\infty \frac{1}{\xi-x}\psi^{[i_0]}(u;dx) = -\mathcal{F}_{i_0}(u;-\xi).$$

Since the function $\tilde{\theta}_{i_0}(u;\xi)$ shall have no poles in $[0,\infty)$, the result follows. □

6. Conclusion

We have presented in this paper a general method for computing the Laplace transform of the transient probability distribution function of the content of a fluid reservoir fed with a source, whose transmission rate is modulated by a general birth and death process. This Laplace transform can be evaluated by solving a polynomial equation (see equation (18)). Once the zeros are known, the quantities $h_i(s)$ for $i = 0, \ldots, i_0$ are computed by solving the system of linear equations (19). These functions then completely determined the two critical functions F_{i_0} and F_{i_0+1}, which are then used for computing the functions F_i for $i > i_0 + 1$ and F_i for $i < i_0$

by using equations (16) and (15), respectively. Moreover, we note that the theory of orthogonal polynomials and continued fractions plays a crucial role in solving the basic equation (6).

The above method can be used for evaluating the Laplace transform of the duration of a busy period of the fluid reservoir as shown in Section 5. The results obtained in this section can be used to study the asymptotic behavior of the busy period when the service rate of the buffer becomes very large. Occupancy periods of the buffer then become rare events and one may expect that buffer characteristics converge to some limits. This will be addressed in further studies.

7. Appendix

A. Proof of Lemma 1

From the recurrence relations (10), the quantities $A_k(s)$ defined by $A_0(s) = 1$ and for $k \geq 1$

$$A_k(s) = |r_0 \dots r_{k-1}| \prod_{j=1}^{k} \alpha_{2j}(s)$$

satisfy the recurrence relation for $k \geq 1$

$$A_{k+1}(s) = (s + \lambda_k + \mu_k)A_k(s) - \lambda_{k-1}\mu_k A_{k-1}(s).$$

It is clear that $A_k(s)$ is a polynomial in variable s. In fact, the polynomials $A_k(s)$ are the successive denominators of the continued fraction

$$\mathcal{G}^e(z) = \cfrac{1}{s + \lambda_0 - \cfrac{\mu_1\lambda_0}{s + \lambda_1 + \mu_1 - \cfrac{\mu_2\lambda_1}{s + \lambda_2 + \mu_2 - \ddots}}}$$

which is itself the even part of the continued fraction

$$\mathcal{G}(s) = \cfrac{\alpha_1}{z + \cfrac{\alpha_2}{1 + \cfrac{\alpha_3}{z + \cfrac{\alpha_4}{1 + \ddots}}}}, \tag{43}$$

where the coefficients α_k are such that $\alpha_1 = 1$, $\alpha_2 = \lambda_0$, and for $k \geq 1$,

$$\alpha_{2k}\alpha_{2k+1} = \lambda_{k-1}\mu_k, \quad \alpha_{2k+1} + \alpha_{2(k+1)} = \lambda_k + \mu_k.$$

It is straightforwardly checked that $\alpha_{2k} = \lambda_{k-1}$ and $\alpha_{2k+1} = \mu_k$ for $k \geq 1$. The continued fraction $\mathcal{G}(s)$ is hence a Stieltjes fraction and is converging for all $s > 0$ if and only if $\sum_{k=0}^{\infty} a_k =$

∞ where the coefficients a_k are defined by

$$\alpha_1 = \frac{1}{a_1}, \quad \alpha_k = \frac{1}{a_{k-1}a_k} \text{ for } k \geq 1.$$

(See Henrici (1977) for details.) It is easily checked that for $k \geq 1$

$$a_{2k} = \frac{1}{\lambda_{k-1}\pi_{k-1}} \quad \text{and} \quad a_{2k+1} = \pi_k.$$

Since the process (Λ_t) is assumed to be ergodic, $\sum_{k\geq 1} a_k = \infty$, which shows that the continued fraction $\mathcal{G}(s)$ is converging for all $s > 0$ and that there exists a unique measure $\varphi(dx)$ such that $\mathcal{G}(s)$ is the Stieltjes transform of $\varphi(dx)$, that is, for all $s \in \mathbb{C} \setminus (-\infty, 0]$

$$\mathcal{G}(s) = \int_0^\infty \frac{1}{z+x}\varphi(dx).$$

The support of $\varphi(dx)$ is included in $[0, \infty)$ and this measure has a mass at point $x_0 \geq 0$ if and only if

$$\sum_{k=0}^{\infty} \frac{A_k(-x_0)^2}{\lambda_0 \dots \lambda_{k-1}\mu_1 \dots \mu_k} < \infty.$$

Since the continued fraction $\mathcal{G}(s)$ is converging for all $s > 0$, we have

$$\sum_{k=0}^{\infty} \frac{A_k(s)^2}{\lambda_0 \dots \lambda_{k-1}\mu_1 \dots \mu_k} = \infty. \tag{44}$$

Since the polynomials $A_k(s)$ are the successive denominator of the fraction $\mathcal{G}^e(s)$, the polynomials $A_k(-s)$, $k \geq 1$, are orthogonal with respect to some orthogonality measure, namely the measure $\varphi(dx)$. From the general theory of orthogonal polynomials Askey (1984); Chihara (1978), we know that the polynomial $A_k(-s)$ has k simple, real, and positive roots. Since the coefficient of the leading term of $A_k(-s)$ is $(-1)^k$, this implies that $A_k(s)$ can be written as $A_k(s) = (s + s_{1,k}) \dots (s + s_{k,k})$ with $s_{i,k} > 0$ for $i = 1, \dots, k$. Hence, $A_k(s) \geq 0$ for all $s \geq 0$ and then, for all $k \geq 0$, $\alpha_k(s) \geq 0$ for all $s \geq 0$ and hence the continued fraction $\mathcal{F}(s, z)$ defined by Equation (9) is a Stieljtes fraction.

The continued fraction $\mathcal{F}(s, z)$ is converging if and only if $\sum_{k=0}^{\infty} a_k(s) = \infty$ where the coefficients $a_k(s)$ are defined by

$$\alpha_1(s) = \frac{1}{a_1(s)}, \quad \alpha_k(s) = \frac{1}{a_{k-1}(s)a_k(s)} \text{ for } k \geq 1.$$

(See Henrici (1977) for details.)

It is easily checked that

$$a_{2k+1}(s) = \frac{|r_k|}{|r_0|} \frac{A_k(s)^2}{\lambda_{k-1} \dots \lambda_0 \mu_k \dots \mu_1} \quad \text{and} \quad a_{2k} = |r_0| \frac{\lambda_0 \dots \lambda_{k-2}\mu_1 \dots \mu_{k-1}}{A_k(s)A_{k-1}(s)}.$$

For $k > i_0$, $r_k \geq r_{i_0+1}$ and then by taking into account Equation (44), we deduce that for all $s > 0$, $\sum_{k=0}^{\infty} a_k(s) = \infty$ and the continued fraction $\mathcal{F}(s;z)$ is then converging for all $s > 0$. For $s = 0$, we have

$$a_{2k}(0) = \frac{|r_0|}{\lambda_{k-1}\pi_{k-1}}$$

and then $\sum_{k=0}^{\infty} a_k(0) = \infty$ since the process (Λ_t) is ergodic (see Condition (2)). This shows that the Stieltjes fraction $\mathcal{F}(s;z)$ is converging for all $s \geq 0$.

B. Selfadjointness properties

We consider in this section the Hilbert space $H_{i_0} = \mathbb{C}^{i_0+1}$ equipped with the scalar product

$$(c,d)_{i_0} = \sum_{k=0}^{i_0} c_k \overline{d_k} |r_k| \pi_k.$$

The main result of this section is the following lemma.

Lemma 6. *For $s \geq 0$, the finite matrix $-R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]} - A_{[i_0]})$ defines a selfadjoint operator in the Hilbert space H_{i_0}; the spectrum is purely point-wise and composed by the (positive) roots of the polynomial $Q_{i_0+1}(s;x)$ defined by Equation* (8)*, denoted by $\zeta_k(s)$ for $k = 0, \ldots, i_0$.*

Proof. The finite matrix $-R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]} - A_{[i_0]})$ is given by

$$\begin{pmatrix} -\frac{s+\lambda_0}{|r_0|} & \frac{\lambda_0}{|r_0|} & 0 & . & . & \\ \frac{\mu_1}{|r_1|} & -\frac{(s+\lambda_1+\mu_1)}{|r_1|} & \frac{\lambda_1}{|r_1|} & . & . & \\ 0 & \frac{\mu_2}{|r_2|} & -\frac{(s+\lambda_2+\mu_2)}{|r_2|} & \frac{\lambda_2}{|r_2|} & . & \\ . & . & . & . & . & \\ & & & & \frac{\mu_{i_0}}{|r_{i_0}|} & -\frac{s+\lambda_{i_0}+\mu_{i_0}}{|r_{i_0}|} \end{pmatrix}.$$

The symmetry of the matrix with respect to the scalar product $(.,.)_{i_0}$ is readily verified by using the relation $\lambda_k \pi_k = \mu_{k+1}\pi_{k+1}$. Since the dimension of the Hilbert space H_{i_0} is finite, the operator associated with the matrix $-R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]} - A_{[i_0]})$ is selfadjoint and its spectrum is purely point-wise.

If f is an eigenvector for the matrix $-R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]} - A_{[i_0]})$ associated with the eigenvalue x, then under the hypothesis that $f_0 = 1$, the sequence f_n verifies the same recurrence relation as $Q_k(s;x)$ for $k = 0, \ldots, i_0 - 1$. This implies that x is an eigenvalue of the above matrix if an only if $Q_{i_0+1}(s;x) = 0$, that is, x is one of the (positive) zeros of the polynomial $Q_{i_0+1}(s;x)$, denoted by $\zeta_k(s)$ for $k = 0, \ldots, i_0$. □

Let us introduce the column vector $Q_{[i_0]}(s, \zeta_k(s))$ for $k = 0, \ldots, i_0$, whose ℓth component is $Q_\ell(s, \zeta_k(s))$. The vector $Q_{[i_0]}(s, \zeta_k(s))$ is the eigenvector associated with the eigenvalue $\zeta_k(s)$ of the operator $-R_{[i_0]}^{-1}(s\mathbb{I}_{[i_0]} - A_{[i_0]})$. From the spectral theorem, the vectors $Q_{[i_0]}(s, \zeta_k(s))$ for

$k = 0, \ldots, i_0$ form an orthogonal basis of the Hilbert space H_{i_0}. The vectors e_j for $j = 0, \ldots, i_0$ such that all entries are equal to 0 except the jth one equal to 1 form the natural orthogonal basis of the space H_{i_0}. We can moreover write for $j = 0, \ldots, i_0$

$$e_j = \sum_{k=0}^{i_0} \alpha_k^{(j)} Q_{[i_0]}(s, \zeta_k(s)).$$

By using the orthogonality of the vectors $Q_{[i_0]}(s, \zeta_k(s))$ for $k = 0, \ldots, i_0$, we have

$$(e_j, Q_{[i_0]}(s, \zeta_k(s)))_{i_0} = |r_j| \pi_j Q_j(s, \zeta_k(s)) = \|Q_{[i_0]}(s, \zeta_k(s))\|_{i_0}^2 \alpha_k^{(j)}$$

where for $f \in H_{i_0}$, $\|f\|_{i_0}^2 = (f, f)_{i_0}$. We hence deduce that

$$|r_j| \pi_j \sum_{k=0}^{i_0} \frac{Q_j(s, \zeta_k(s)) Q_\ell(s, \zeta_k(s))}{\|Q_{[i_0]}(s, \zeta_k(s))\|_{i_0}^2} = \delta_{j,\ell},$$

where $\delta_{j,\ell}$ is the Kronecker symbol. It follows that if we define the measure $\psi_{[i_0]}(s; dx)$ by

$$\psi_{[i_0]}(s; dx) = |r_0| \sum_{k=0}^{i_0} \frac{1}{\|Q_{[i_0]}(s, \zeta_k(s))\|_{i_0}^2} \delta_{\zeta_k(s)}(dx) \tag{45}$$

the polynomials $Q_k(s, x)$ for $k = 0, \ldots, i_0$ are orthogonal with respect to the above measure, that is, they verify

$$\int_0^\infty Q_j(s, x) Q_\ell(s, x) \psi_{[i_0]}(s; dx) = \frac{|r_0|}{|r_j| \pi_j} \delta_{j,\ell},$$

and the total mass of the measure $\psi_{[i_0]}(s; dx)$ is equal to 1, i.e,

$$\int_0^\infty \psi_{[i_0]}(s; dx) = 1.$$

8. References

Adan, I. & Resing, J. (1996). Simple analysis of a fluid queue driven by an M/M/1 queue, *Queueing Systems - Theory and Applications*, Vol. 22, pp. 171–174.

Aggarwal, V., Gautam, N., Kumara, S. R. T. & Greaves, M. (2005). Stochastic fluid flow models for determining optimal switching thresholds, *Performance Evaluation*, Vol. 59, pp. 19–46.

Ahn, S. & Ramaswami, V. (2003). Fluid flow models and queues - a connection by stochastic coupling, *Stochastic Models*, Vol. 19, No. 3, pp. 325–348.

Ahn, S. & Ramaswami, V. (2004). Transient analysis of fluid flow models via stochastic coupling to a queue, *Stochastic Models*, Vol. 20, No. 1, pp. 71–101.

Anick, D., Mitra, D. & Sondhi, M. M. (1982). Stochastic theory of a data-handling system with multiple sources, *Bell System Tech. J.*, Vol. 61, No. 8, pp. 1871–1894.

Askey, R. & Ismail, M. (1984). Recurrence relations, continued fractions, and orthogonal polynomials, *Memoirs of the American Mathematical Society*, Vol. 49, No. 300.

Asmussen, S. (1987). Applied probability and queues, *J. Wiley and Sons*.

Badescu, A., Breuer, L., da Silva Soares, A., Latouche, G., Remiche, M.-A. & Stanford, D. (2005). Risk processes analyzed as fluid queues, *Scandinav. Actuar. J.*, Vol. 2, pp. 127–141.

Barbot, N., Sericola, B. & Telek, M. (2001). Dsitribution of busy period in stochastic fluid models, *Stochastic Models*, Vol. 17, No. 4, pp. 407–427.

Barbot, N. & Sericola, B. (2002). Stationary solution to the fluid queue fed by an M/M/1 queue, *Journ. Appl. Probab.*, Vol. 39, pp. 359–369.

Chihara, T. S. (1978). *An introduction to orthogonal polynomials*. Gordon and Breach, New York, 1978.

da Silva Soares, A. & Latouche, G. (2002). Further results on the similarity between fluid queues and QBDs, In G. Latouche and P. Taylor, editors, *Proc. of the 4th Int. Conf. on Matrix-Analytic Methods (MAM'4)*, Adelaide, Australia, 89–106, World Scientific.

da Silva Soares, A. & Latouche, G. (2006). Matrix-analytic methods for fluid queues with finite buffers, *Performance Evaluation*, Vol. 63, No. 4, pp. 295–314.

Guillemin, F. (2012). Spectral theory of birth and death processes, *Submitted for publication*.

Guillemin, F. & Pinchon, D. (1999). Excursions of birth and death processes, orthogonal polynomials, and continued fractions, *J. Appl. Prob.*, Vol. 36, pp. 752–770.

Guillemin, F. & Sericola, B. (2007). Stationary analysis of a fluid queue driven by some countable state space Markov chain, *Methodology and Computing in Applied Probability*, Vol. 9, pp. 521–540.

Henrici, P. (1977). Applied and computational complex analysis, *Wiley, New York*, Vol. 2.

Igelnik, B., Kogan, Y., Kriman, V. & Mitra, D. (1995). A new computational approach for stochastic fluid models of multiplexers with heterogeneous sources, *Queueing Systems - Theory and Applications*, Vol. 20, pp. 85–116.

Kleinrock, L. (1975). Queueing Systems, *J. Wiley*, Vol. 1.

Kobayashi, H. & Ren, Q. (1992). A mathematical theory for transient analysis of communication networks, *IEICE Trans. Communications*, Vol. 75, No. 12, pp. 1266–1276.

Kosten, L. (1984). Stochastic theory of data-handling systems with groups of multiple sources, In *Proceedings of the IFIP WG 7.3/TC 6 Second International Symposium on the Performance of Computer-Communication Systems*, Zurich, Switzerland, pp. 321–331.

Kumar, R., Liu, Y. & Ross, K. W. (2007). Stochastic Fluid Theory for P2P Streaming Systems, In *Proceedings of INFOCOM*, Anchorage, Alaska, USA, pp. 919–927.

Mitra, D. (1987). Stochastic fluid models, In *Proceedings of Performance'87, P. J. Courtois and G. Latouche Editors*, Brussels, Belgium, pp. 39–51.

Mitra, D. (1988). Stochastic theory of a fluid model of producers and consumers coupled by a buffer, *Advances in Applied Probability*, Vol. 20, pp. 646–676.

Nabli, H. & Sericola, B. (1996). Performability analysis: a new algorithm, *IEEE Trans. Computers*, Vol. 45, pp. 491–494.

Nabli, H. (2004). Asymptotic solution of stochastic fluid models, *Performance Evaluation*, Vol. 57, pp. 121–140.

Parthasarathy, P. R., Vijayashree, K. V. & Lenin, R. B. (2004). Fluid queues driven by a birth and death process with alternating flow rates, *Mathematical Problems in Engineering*, Vol. 5, pp. 469–489.

Ramaswami, V. (1999). Matrix analytic methods for stochastic fluid flows, In D. Smith and P. Hey, editors, *Proceedings of the 16th International Teletraffic Congress : Teletraffic Engineering in a Competitive World (ITC'16)*, Edinburgh, UK, Elsevier, pp. 1019–1030.

Ren, Q. & Kobayashi, H. (1995). Transient solutions for the buffer behavior in statistical multiplexing, *Performance Evaluation*, Vol. 23, pp. 65–87.

Rogers, L. C. G. (1994). Fluid models in queueing theory and wiener-hopf factorization of Markov chains, *Advances in Applied Probability*, Vol. 4, No. 2.

Rogers, L. C. G. & Shi, Z. (1994). Computing the invariant law of a fluid model, *Journal of Applied Probability*, Vol. 31, No. 4, pp. 885–896.

Sericola, B. (1998). Transient analysis of stochastic fluid models, *Performance Evaluation*, Vol. 32, pp. 245–263.

Sericola, B. & Tuffin, B. (1999). A fluid queue driven by a Markovian queue, *Queueing Systems - Theory and Applications*, Vol. 31, pp. 253–264.

Sericola, B. (2001). A finite buffer fluid queue driven by a Markovian queue, *Queueing Systems - Theory and Applications*, Vol. 38, pp. 213–220.

Sericola, B., Parthasarathy, P. R. & Vijayashree, K. V. (2005). Exact transient solution of an M/M/1 driven fluid queue, *Int. Journ. of Computer Mathematics*, Vol. 82, No. 6.

Stern, T. E. & Elwalid, A. I. (1991). Analysis of separable Markov-modulated rate models for information-handling systems, *Advances in Applied Probability*, Vol. 23, pp. 105–139.

Tanaka, T., Hashida, O. & Takahashi, Y. (1995). Transient analysis of fluid models for ATM statistical multiplexer, *Performance Evaluation*, Vol. 23, pp. 145–162.

van Dorn, E. A. & Scheinhardt, W. R. (1997). A fluid queue driven by an infinite-state birth and death process, In V. Ramaswami and P. E. Wirth, editors, *Proceedings of the 15th International Teletraffic Congress : Teletraffic Contribution for the Information Age (ITC'15)*, Washington D.C., USA, Elsevier, pp. 465–475.

vanForeest, N., Mandjes, M. & Scheinhardt, W. R. (2003). Analysis of a feedback fluid model for TCP with heterogeneous sources, *Stochastic Models*, Vol. 19, pp. 299–324.

Virtamo, J. & Norros, I. (1994). Fluid queue driven by an M/M/1 queue, *Queueing Systems - Theory and Applications*, Vol. 16, pp. 373–386.

17

Optimal Control Strategies for Multipath Routing: From Load Balancing to Bottleneck Link Management

C. Bruni, F. Delli Priscoli, G. Koch, A. Pietrabissa and L. Pimpinella
Dipartimento di Informatica e Sistemistica "A. Ruberti",
"Sapienza" Università di Roma, Roma,
Italy

1. Introduction

In this work we face the Routing problem defined as an optimal control problem, with control variables representing the percentages of each flow routed along the available paths, and with a cost function which accounts for the distribution of traffic flows across the network resources (multipath routing). In particular, the scenario includes the load balancing problem already dealt with in a previous work (Bruni et al., 2010) as well as the bottleneck minimax control problem. The proposed approaches are then compared by evaluating the performances of a sample network.

In a given network, the resource management problem consists in taking decisions about handling the traffic amount which is carried by the network, while respecting a set of Quality of Service (QoS) constraints.

As stated in Bruni et al., 2009a, b, the resource management problem is hardly tackled by a single procedure. Rather, it is currently decomposed in a number of subproblems (Connection Admission Control (CAC), traffic policing, routing, dynamic capacity assignment, congestion control, scheduling), each one coping with a specific aspect of such problem. In this respect, the present work is embedded within the general approach already proposed by the authors in Bruni et al., 2009a, b, according to which each of the various sub-problems is given a separate formulation and solution procedure, which strives to make the other sub-problems easier to be solved. More specifically, the above mentioned approach consists in charging the CAC with the task of deciding, on the basis of the network congestion state, new connection admission/blocking and possible forced dropping of the in-progress connections with the aim of maximizing the number of accepted connections, whilst satisfying the QoS requirements.

According to the proposed approach, the role of the other resource management procedures is the one of keeping the network as far as possible far from the congestion state. Indeed, the more the network is kept far from congestion, the higher is the number of new connection set-up attempts that can be accepted by the CAC without infringing the QoS constraints,

and hence the traffic carried by the network increases. By so doing, the CAC and the other resource management procedures can work in a consistent way, while being kept independent.

This work deals with the multipath routing problem. Multipath routing is a widespread topic in the literature. For example, Cidon et al., 1999, and Banner and Orda, 2007, demonstrate the advantages of multipath routing with respect to single-path routing in terms of network performances; Chen et al., 2004, considers the multipath routing problem under bandwidth and delay constraints; Lin and Shroff, 2006, formulate the multipath routing problem as a utility maximization problem with bandwidth constraints; Guven et al., 2008, extend the multipath routing to multicast flows; Jaffe, 1981, Tsai et al., 2006, Tsai and Kim, 1999 deal with the multipath routing as a minimax optimization problem.

In this work we face the multipath routing problem formulated as an optimal control problem, with control variables representing the percentages of each flow routed along the available paths. As a matter of fact, in the most advanced networks each flow can be simultaneously routed over more than one path: the routing procedure has to decide the percentages of the traffic belonging to the considered flow which have to be routed over the paths associated to the flow in question. According to the above mentioned vision, we assume that other resource management control units (specifically the CAC) already dealt with and decided about issues such as how many, which ones, when and for how long connections have to be admitted in the network, with specific QoS constraints (related to losses and delays) to be satisfied. Therefore, the routing control unit has to deal with an already defined offered traffic. Thus, the admissible set for the routing control variables turns out to be closed, bounded and non-empty, and the existence of (at least) an optimal solution of the routing problem is guaranteed.

The goal of an optimal routing policy aims the routing problem solution towards a network traffic pattern which should make QoS requirements and consequently the CAC task (implicitly) easier to be satisfied. The quality of the routing solution will be evaluated by different performance indices, which take a nominal capacity for each link into account.

As far as the dynamical aspects of a routing problem, we first note that explicitly accounting for them would call for a reliable and sufficiently general dynamical model for the offered traffic. However it is widely acknowledged that such a model is not available and hard to design, due to unpredictable features of Internet traffic. And, in any case, the requested dynamical characters are committed to the CAC procedures, where the more reliable connection dynamics model along with the feedback structure may properly handle the issue.

In addition, a non-dynamical set up for the routing problem makes it much easier to be dealt with. Moreover, this approach could be justified by assuming that the time scale for changes in the routing policy is surely slower than the bit rate fluctuations in the in-progress connections, but it is reasonably faster than the evolution of traffic statistical features. Thus, the routing policy has to be periodically computed to fit the most likely traffic pattern at each given period of time.

In this work, we consider the possibility/opportunity of splitting the given network into sub-networks as detailed in Bruni et al., 2010 each one controlled by a separate subset of variables.

This work is organized as follows. In Section 2, a definition for a reference communication network and its decomposition is given, which is useful for the routing problem; in Sections 3, we in depth study the optimal routing control problem with reference to a number of different cost functions; Section 4 shows some results in order to evaluate the performance and to compare the found optimal solutions for traffic balancing and bottleneck link management; finally, concluding remarks in Section 5 end the work.

2. Reference telecommunication network definition and decomposition

At any fixed time, the telecommunication network can be defined in terms of its topological description as well as in terms of its traffic pattern. As far as network topology is concerned, we consider the network nodes $n \in \mathrm{N} = \{n_1,n_2,\ldots,n_N\}$ and the network links defined as ordered pairs of nodes $l \in \Lambda = \{l_1,l_2,\ldots,l_L\}$. To describe the network traffic request we first define a path $v \in \Omega = \{v_1,v_2,\ldots,v_V\}$ as a collection of consecutive links, denoted by Λ_v, from an ingoing node i to an outgoing node j (where $i,j \in \mathbf{N}$). Moreover a certain set of different Service Classes $k \in \mathrm{K} = \{k_1,k_2,\ldots,k_K\}$, is defined, each one characterized by a set of Quality of Service (QoS) parameters. According to the most recent trends, the QoS control is performed on a per flow basis, where a flow $f \in \Phi = \{f_1,f_2,\ldots,f_F\}$ is defined as the triple $f = (n_i,n_j,k_p)$, with n_i denoting the ingoing node, n_j denoting the outgoing node and k_p denoting the service class. The traffic associated with a given flow f may possibly be routed on a set Ω_f of one or more paths. We further introduce the set of indices $\{a(l,v),\ l \in \Lambda,\ v \in \Omega\}$, defined as follows:

$$a(l,v) = \begin{cases} 1, \text{ if } l \in v \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

For each link $l \in \Lambda$, at the given time, we may consider its occupancy level $c(l)$ defined as the sum of all contributions to the occupancy due to the flows routed on the link itself. Each contribution of this type will be quantified by the bit rate $R(l,f)$ which, in turn, is the sum of bit rates of all in-progress connections going through the link l and relevant to the flow f, possibly weighted by a coefficient $\alpha(l,f)$ which accounts for the specific need of the flow itself. Therefore we have:

$$c(l) = \sum_{f \in \Phi} \alpha(l,f) R(l,f) \tag{2}$$

where $\alpha(l,f)$ are positive known coefficients which take into account the fact that some technologies differentiate the classes of service by varying modulation, coding, and so on. For each link l, we consider the so-called nominal capacity $c_{NOM}(l)$, that is the value of the occupancy level suggested for a proper behaviour of the link (typically in terms of QoS)[1].

[1] $c(l)$ and $c_{NOM}(l)$ can be interpreted as generalizations of "load factor" and "Noise Rise" in UMTS (see Holma and Toskala, 2002).

that indicates the fraction of $R(f)$ to be routed on path $v \in \Omega_f$. Then, due to the bit conservation law, we have:

$$R(l,f) = \sum_{v \in \Omega_f} \alpha(l,f) R(f) u(f,v) \tag{3}$$

where obviously:

$$\begin{aligned} & u(f,v) \in [0,1], \ \forall f \in \Phi, \ \forall v \in \Omega_f \\ & \sum_{v \in \Omega_f} u(f,v) = 1, \ \forall f \in \Phi \end{aligned} \tag{4}$$

As shown in Bruni *et al.*, 2010, with reference to the routing control problem, the link set Λ might be decomposed into separated subclasses $\Lambda^{(j)}$, $j = 0,1,2,\ldots,P$, each of them involving separate subsets of control variables, where $\Lambda^{(0)}$ is the set, possibly empty, of links that cannot be controlled by any control variable and which therefore they are not involved in any routing control problem.

For every communicating class of links $\Lambda^{(j)} \subset \Lambda$, there exists the (uniquely) corresponding communicating class of flows $\Phi^{(j)} \subset \Phi$ defined as the set of flows such that, for each $f \in \Phi^{(j)}$, there exists (at least) a link $l \in \Lambda^{(j)}$, and therefore a pair of links (generally depending on f itself), which are controllable with respect to f. Clearly, the set $\Phi^{(0)}$ coincides with the empty set. We now observe that the set $\{\Phi^{(j)}, \ j \geq 0\}$ of flow communicating classes forms a partition of Φ, corresponding to the fact that the set $\{\Lambda^{(j)}, \ j \geq 0\}$ of link communicating classes forms a partition of Λ. This partition for Λ and Φ immediately induces a partition of the network. Note that each j-th part of the network is controlled by a corresponding subvector of control variables, later defined as $u^{(j)}$ independently of the other parts; the components of the vector $u^{(j)}$ are the variables $u(f,v)$, $f \in \Phi^{(j)}$, $v \in \Omega_f$. In the following $\{\Lambda^{(j)}, \Phi^{(j)}\}$ will denote a sub-network. We will use the detailed network decomposition procedure described in Bruni *et al.*, 2010, facing the routing control problem in each sub-network (but in $\Lambda^{(0)}$).

3. A rationale for the network loading

In the following, we will focus attention on the routing problem for any given sub-network $\{\Lambda^{(j)}, \Phi^{(j)}\}$. As mentioned above, any such problem is characterized by a set $u^{(j)}$ of control variables, which may be (optimally) selected independently of the other ones. As stated in Bruni *et al.*, 2010, the admissible set for $u^{(j)}$ is defined by the constraints:

$$u(f,v) \in [0,1], \ \forall f \in \Phi^{(j)}, \ \forall v \in \Omega_f \tag{5}$$

$$\sum_{v \in \Omega_f} u(f,v) = 1, \ \forall f \in \Phi^{(j)} \tag{6}$$

so that the set itself turns out to be convex. From here on, for sake of simplicity the apices j will be dropped.

The optimal choice for u within its (convex) admissible set may be performed according to a cost function which assesses the network loading. In a previous work Bruni *et al.*, 2010, the control goal was the normalized load balancing in the sub-network, evaluated by the function:

$$J(u) = \sum_{l \in \Lambda} \left(\frac{c(l)}{c_{NOM}(l)} - k \right)^2 \tag{7}$$

with k a given constant. If, for any given u, we optimize (7) with respect to k, we get:

$$k = \frac{1}{L} \sum_{l \in \Lambda} \frac{c(l)}{c_{NOM}(l)} \tag{8}$$

with L denoting the cardinality of Λ. In Bruni *et al.*, 2010, and Bruni *et al.*, 2010 (to appear), a shortcoming of (7) was enlightened, which is due to the partial controllability property (therein defined) of some of the links. These links, in the following referred to as "ballast", are such that they are bound to accept traffic flows not controlled by the components of the control vector u. Thus other choices of the cost function might be considered which more explicitly account for the network overloading.

One first possibility is to assess the link overflow setting $k = 0$ in (7), thus more generally arriving at the functions:

$$J(u) = \sum_{l \in \Lambda} \left(\frac{c(l)}{c_{NOM}(l)} \right)^m \tag{9}$$

for some integer $m \geq 1$. If the target is to give more importance to the links belonging to several paths the function (9) can be rewritten as follows:

$$J(u) = \sum_{v \in \Omega} \sum_{l \in \Lambda_v} \left(\frac{c(l)}{c_{NOM}(l)} \right)^m \tag{10}$$

According to (9), (10) we try to distribute the total load in the network in such a way that the higher the normalized load for a link is, the stronger is the effort in reducing it. This selective attention to the most heavily loaded links progressively increases with m. As m keeps increasing, then function (10) is approximated by:

$$J(u) = \sum_{v \in \Omega} (G_v)^m \tag{11}$$

where:

$$G_v = \max_{l \in \Lambda_v} \frac{c(l)}{c_{NOM}(l)} \tag{12}$$

Thus for each path v the optimization attention is just focused on the most heavily loaded link of the path itself (bottleneck). Eventually we can consider the worst bottleneck load over the whole sub-network:

$$J(u) = \max_{v \in \Omega} G_v \tag{13}$$

Remark. Some methods are proposed in the literature to solve the above minimax optimization problem (see Warren *et al.*, 1967, Osborne and Wetson, 1969, Blander *et al.*, 1972, Blander and Charambous, 1972). The original minimax problem (11) is equivalent to the following:

$$\min_{u,g \in U} J(u,g) \tag{14}$$

$$J(u,g) = \sum_{v \in \Omega} [g(v)]^m \tag{15}$$

$$U = \left\{ (u,g) \in R^{V(F+1)} : u(f,v) \geq 0, \sum_{v \in \Omega_f} u(f,v) = 1, \right.$$
$$\left. \frac{c(l)}{c_{NOM}(l)} \leq g(v), \forall f \in \Phi, \forall v \in \Omega_f \right\} \tag{16}$$

where g is the vector of auxiliary variables $g(v)$, $v \in \Omega$. This is a nonlinear (linear if $m = 1$, quadratic if $m = 2$) programming problem that can be solved by well-established methods. We observe that the equivalence lies in the fact that, once (14) (15) (16) is solved, the optimal value assumed by $g(v)$ coincides with G_v in (12), for $v \in \Omega$, i.e., it represents the normalized bottleneck link load of path v.

The load balancing problem (7) (8), with constraints (5) (6) and the bottleneck load management problem (14) (15) (16) are easily seen to be convex. This allows standard minimization routines to be used for its solution, such as MatLab simulation tools.

Remark. The cost function (13) enlightens a further advantage of network decomposition. Indeed, in case the decomposition had not been performed, then (13) would describe an ill-posed optimal control problem whenever the worst bottleneck over the whole network happens to be an uncontrollable link. Similar considerations hold for cost function (11).

4. Evaluation and comparison of optimal routing procedures

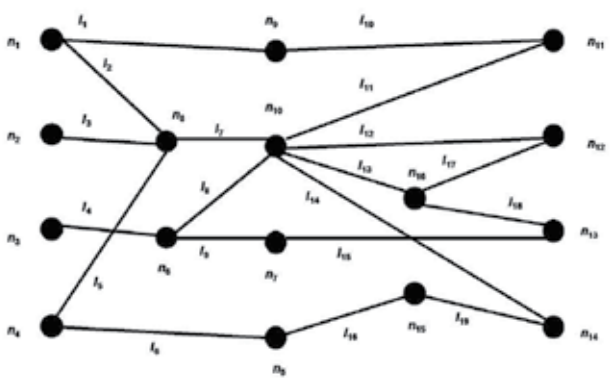
4.1 Network structure and decomposition

The considered scenario is composed by 16 nodes and 19 links (see Fig. 1 a)). The traffic pattern involves 4 traffic flows of the same service class k, from 4 source nodes n_i, $i = 1,..,4$, to 4 different destination nodes n_j, $j = 11,..,14$. The traffic pattern is described by the set of traffic flows $\Phi = \{f_1,f_2,f_3,f_4\}$, where each traffic flow is identified by the following triples: $f_1 = (n_1,n_{11},k)$, $f_2 = (n_2,n_{12},k)$, $f_3 = (n_3,n_{13},k)$, $f_4 = (n_4,n_{14},k)$. After performing the network decomposition as in Bruni *et al.*, 2010, we recognize three sub-networks (see, Fig. 1 b), c) and d)). The network topology is summarized in Table 1, where the network decomposition is reported as well.

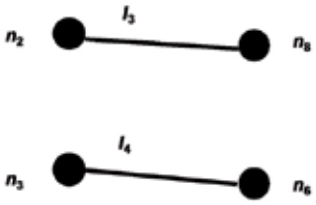
	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}	l_{11}	l_{12}	l_{13}	l_{14}	l_{15}	l_{16}	l_{17}	l_{18}	l_{19}	f_1	f_2	f_3	f_4
c_{NOM} [kbps]	10	10	5.4	5.4	5.4	5.4	10	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4				
v_1	x									x										x			
v_2		x					x				x									x			
v_3			x				x					x									x		
v_4			x				x						x				x				x		
v_5				x				x					x					x				x	
v_6				x					x						x							x	
v_7					x		x							x									x
v_8						x										x		x					x
$\Lambda^{(0)}$			x	x																			
$\Lambda^{(1)}$	x	x			x	x	◊			x	x			x		x			x	x			x
$\Lambda^{(2)}$								x	x			x	x		x		x	x			x	x	

Table 1. Network Topology and Decomposition; the first row shows the nominal link capacities in [Mbps]; the generic entry (l_i,v_j) is denoted by 'x' if $l_i \in \Lambda^{(j)}{}_v$; the generic entry (f_i,v_j) is denoted by 'x' if it is possible to route f_i on path v_j; the generic entry $(l_i,\Lambda^{(j)})$ is denoted by 'x' if $l_i \in \Lambda^{(j)}$, or by '◊' if $l_i \in \Lambda^{(j)}$ and l_i is a ballast link; the generic entry $(f_i,\Lambda^{(j)})$ is denoted by 'x' if $f_i \in \Phi^{(j)}$.

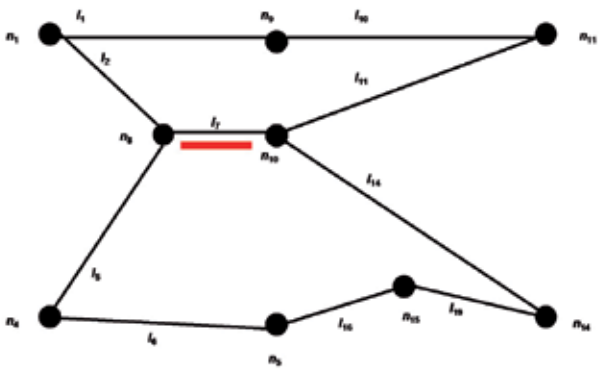
The considered scenario has been simulated with MatLab. In particular we have tested two simuation sets reported in subsection 4.2 and 4.3respectively. In subsection 4.2 we considered the Bottleneck Link Management by varying the weights of the bottleneck loads, while in subsection 4.3 we made comparisons between Load Balancing and Bottleneck Link Management.



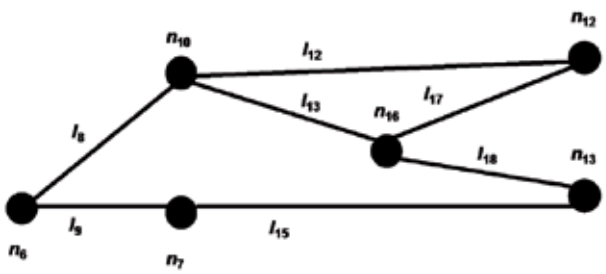
a)



b)



c)

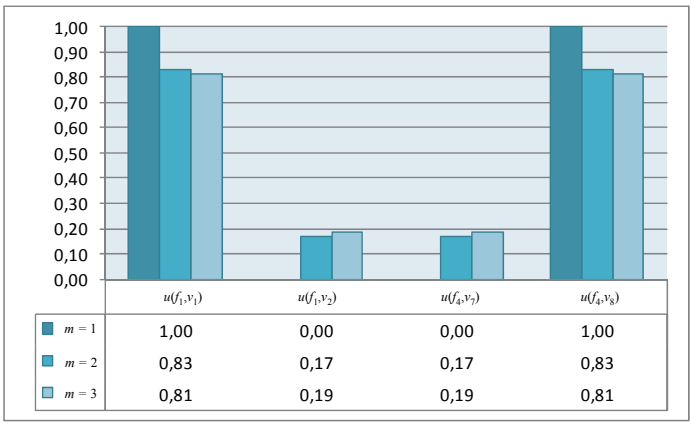


d)

Fig. 1. a) Global Network, b) Sub-network 0 ($\Lambda^{(0)}$), c) Sub-network 1 ($\Lambda^{(1)}$), d) Sub-network 2 ($\Lambda^{(2)}$).

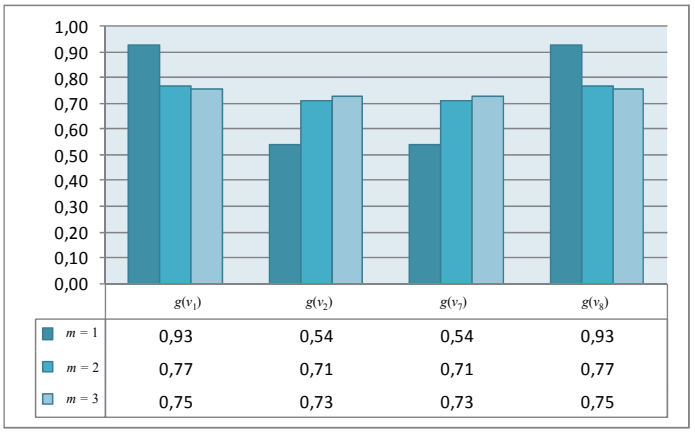
4.2 Optimal routing for different weights of bottleneck loads

In this simulation set we consider that the bit rate of traffic flows f_1, f_3, f_4 is equal to 5 Mbps whilst the bit rate of traffic flow f_2 is equal to 5.4 Mbps. Fig. 2 and 3 show the dependence of the optimal solutions on index m of the Bottleneck Link Management problem.



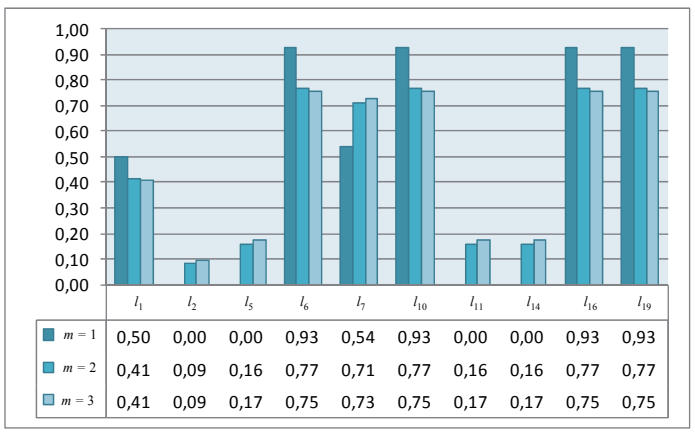
	$u(f_1,v_1)$	$u(f_1,v_2)$	$u(f_4,v_7)$	$u(f_4,v_8)$
$m = 1$	1,00	0,00	0,00	1,00
$m = 2$	0,83	0,17	0,17	0,83
$m = 3$	0,81	0,19	0,19	0,81

a)



	$g(v_1)$	$g(v_2)$	$g(v_7)$	$g(v_8)$
$m = 1$	0,93	0,54	0,54	0,93
$m = 2$	0,77	0,71	0,71	0,77
$m = 3$	0,75	0,73	0,73	0,75

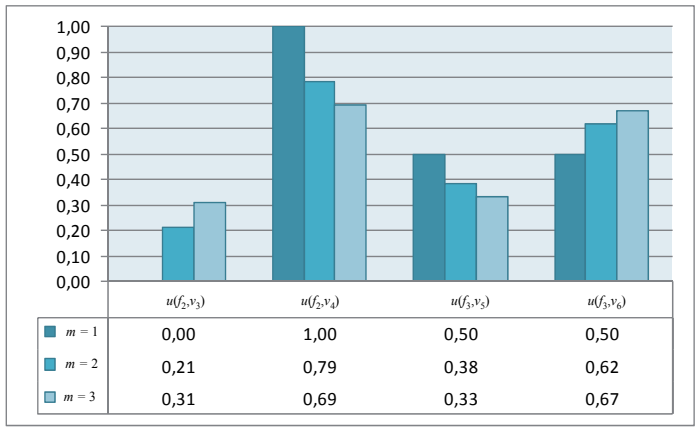
b)



	l_1	l_2	l_5	l_6	l_7	l_{10}	l_{11}	l_{14}	l_{16}	l_{19}
$m = 1$	0,50	0,00	0,00	0,93	0,54	0,93	0,00	0,00	0,93	0,93
$m = 2$	0,41	0,09	0,16	0,77	0,71	0,77	0,16	0,16	0,77	0,77
$m = 3$	0,41	0,09	0,17	0,75	0,73	0,75	0,17	0,17	0,75	0,75

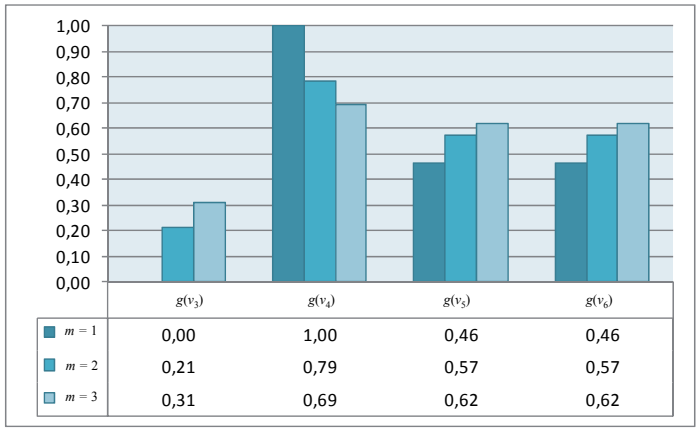
c)

Fig. 2. Sub-network 1: a) optimal control variables, b) bottleneck link loads, c) link loads.



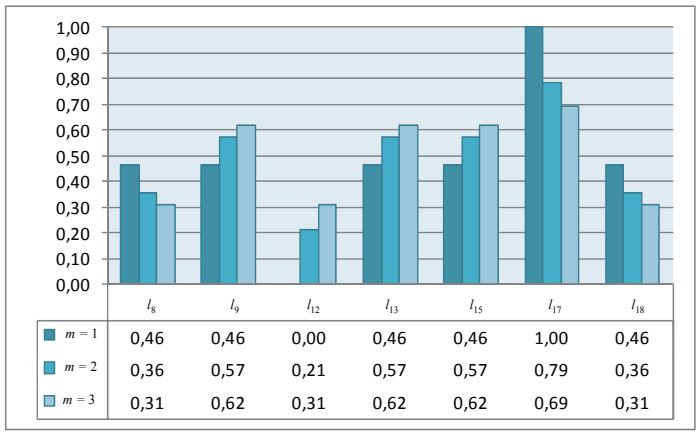
	$u(f_2,v_3)$	$u(f_2,v_4)$	$u(f_3,v_5)$	$u(f_3,v_6)$
$m = 1$	0,00	1,00	0,50	0,50
$m = 2$	0,21	0,79	0,38	0,62
$m = 3$	0,31	0,69	0,33	0,67

a)



	$g(v_3)$	$g(v_4)$	$g(v_5)$	$g(v_6)$
$m = 1$	0,00	1,00	0,46	0,46
$m = 2$	0,21	0,79	0,57	0,57
$m = 3$	0,31	0,69	0,62	0,62

b)



	l_8	l_9	l_{12}	l_{13}	l_{15}	l_{17}	l_{18}
$m = 1$	0,46	0,46	0,00	0,46	0,46	1,00	0,46
$m = 2$	0,36	0,57	0,21	0,57	0,57	0,79	0,36
$m = 3$	0,31	0,62	0,31	0,62	0,62	0,69	0,31

c)

Fig. 3. Sub-network 2: a) optimal control variables, b) bottleneck link loads, c) link loads.

4.3 Comparisons between load balancing and bottleneck link management

In this simulation set we consider that all the traffic sources transmit with an increasing trend from 4.5 Mbps to 8.5 Mbps. Tables 2-5 show the network load as the sources bit rate increase, and compares the optimal bottleneck control solutions for m = 1, 2, 3 with the load balancing optimal solution.

In Tables 2-5, we denote by bold characters the normalized link loads exceeding 1; hereinafter, the corresponding links will be denoted as overloaded links.

The bottleneck control for $m \geq 2$ manages a higher network load than the load balancing approach. In fact, the tables show that the solutions of the bottleneck control problem are such that no link is overloaded until the flow rates exceed 5 Mbps, 6.5 Mbps and 6.5 Mbps for m = 1,2,3, respectively; on the other hand, the load balancing solutions are such that no link is overloaded until the flow rates exceed 5 Mbps. Similar results are obtained for sub-network 2.

Rate [Mbps]	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5
$u(f_1,v_1)$	1,00									
$u(f_1, v_2)$	0,00									
$u(f_4, v_7)$	0,00									
$u(f_4, v_8)$	1,00									
$g(v_1)$	0,74	0,83	0,93	**1,02**	**1,11**	**1,20**	**1,30**	**1,39**	**1,48**	**1,57**
$g(v_2)$	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85
$g(v_7)$	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85
$g(v_8)$	0,74	0,83	0,93	**1,02**	**1,11**	**1,20**	**1,30**	**1,39**	**1,48**	**1,57**
l_1	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85
l_2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
l_5	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
l_6	0,74	0,83	0,93	**1,02**	**1,11**	**1,20**	**1,30**	**1,39**	**1,48**	**1,57**
l_7	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85
l_{10}	0,74	0,83	0,93	**1,02**	**1,11**	**1,20**	**1,30**	**1,39**	**1,48**	**1,57**
l_{11}	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
l_{14}	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
l_{16}	0,74	0,83	0,93	**1,02**	**1,11**	**1,20**	**1,30**	**1,39**	**1,48**	**1,57**
l_{19}	0,74	0,83	0,93	**1,02**	**1,11**	**1,20**	**1,30**	**1,39**	**1,48**	**1,57**

Table 2. Sub-network 1: Optimal Solutions under bottleneck control, m =1.

Rate [Mbps]	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5
$u(f_1,v_1)$	0,81									
$u(f_1, v_2)$	0,19									
$u(f_4, v_7)$	0,19									
$u(f_4, v_8)$	0,81									
$g(v_1)$	0,60	0,67	0,75	0,82	0,90	0,97	1,05	1,12	1,20	1,27
$g(v_2)$	0,55	0,62	0,69	0,76	0,83	0,90	0,97	1,04	1,11	1,18
$g(v_7)$	0,55	0,62	0,69	0,76	0,83	0,90	0,97	1,04	1,11	1,18
$g(v_8)$	0,60	0,67	0,75	0,82	0,90	0,97	1,05	1,12	1,20	1,27
l_1	0,32	0,36	0,40	0,44	0,48	0,52	0,57	0,61	0,65	0,69
l_2	0,08	0,09	0,10	0,11	0,12	0,13	0,13	0,14	0,15	0,16
l_5	0,14	0,16	0,18	0,20	0,21	0,23	0,25	0,27	0,29	0,30
l_6	0,60	0,67	0,75	0,82	0,90	0,97	**1,05**	**1,12**	**1,20**	**1,27**
l_7	0,55	0,62	0,69	0,76	0,83	0,90	0,97	**1,04**	**1,11**	**1,18**
l_{10}	0,60	0,67	0,75	0,82	0,90	0,97	**1,05**	**1,12**	**1,20**	**1,27**
l_{11}	0,14	0,16	0,18	0,20	0,21	0,23	0,25	0,27	0,29	0,30
l_{14}	0,14	0,16	0,18	0,20	0,21	0,23	0,25	0,27	0,29	0,30
l_{16}	0,60	0,67	0,75	0,82	0,90	0,97	**1,05**	**1,12**	**1,20**	**1,27**
l_{19}	0,60	0,67	0,75	0,82	0,90	0,97	**1,05**	**1,12**	**1,20**	**1,27**

Table 3. Sub-network 1: Optimal Solutions under bottleneck control, m =2.

Rate [Mbps]	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5
$u(f_1,v_1)$	0,79									
$u(f_1, v_2)$	0,21									
$u(f_4, v_7)$	0,21									
$u(f_4, v_8)$	0,79									
$g(v_1)$	0,59	0,66	0,73	0,81	0,88	0,95	1,03	1,10	1,18	1,25
$g(v_2)$	0,57	0,64	0,71	0,78	0,85	0,92	0,99	1,06	1,13	1,20
$g(v_7)$	0,57	0,64	0,71	0,78	0,85	0,92	0,99	1,06	1,13	1,20
$g(v_8)$	0,59	0,66	0,73	0,81	0,88	0,95	1,03	1,10	1,18	1,25
l_1	0,32	0,36	0,40	0,44	0,48	0,52	0,56	0,59	0,63	0,67
l_2	0,08	0,09	0,10	0,11	0,12	0,13	0,14	0,16	0,17	0,18
l_5	0,15	0,17	0,19	0,21	0,23	0,25	0,27	0,29	0,31	0,33
l_6	0,59	0,66	0,73	0,81	0,88	0,95	**1,03**	**1,10**	**1,18**	**1,25**
l_7	0,57	0,64	0,71	0,78	0,85	0,92	0,99	**1,06**	**1,13**	**1,20**
l_{10}	0,59	0,66	0,73	0,81	0,88	0,95	**1,03**	**1,10**	**1,18**	**1,25**
l_{11}	0,15	0,17	0,19	0,21	0,23	0,25	0,27	0,29	0,31	0,33
l_{14}	0,15	0,17	0,19	0,21	0,23	0,25	0,27	0,29	0,31	0,33
l_{16}	0,59	0,66	0,73	0,81	0,88	0,95	**1,03**	**1,10**	**1,18**	**1,25**
l_{19}	0,59	0,66	0,73	0,81	0,88	0,95	**1,03**	**1,10**	**1,18**	**1,25**

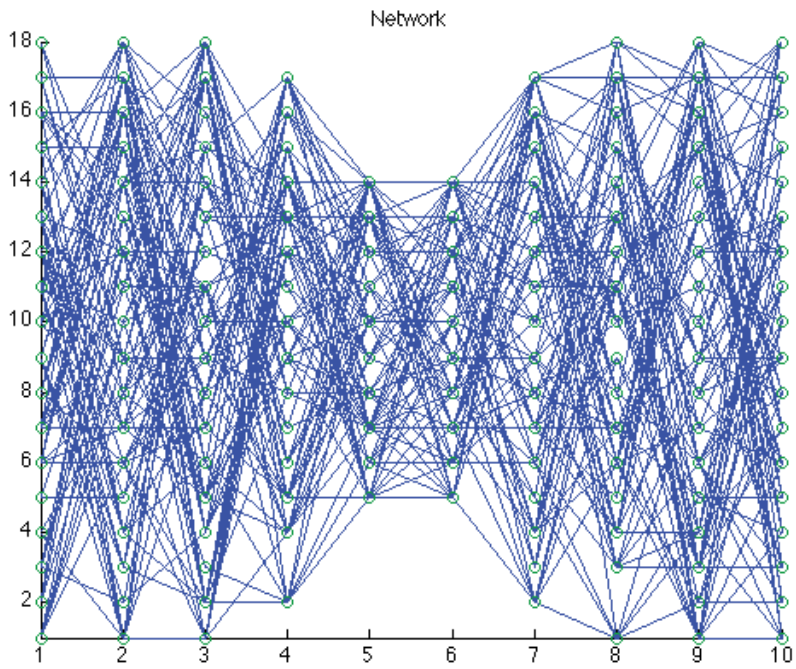
Table 4. Sub-network 1: Optimal Solutions under bottleneck control, m =3.

Rate [Mbps]	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5
$u(f_1,v_1)$	0,60									
$u(f_1, v_2)$	0,40									
$u(f_4, v_7)$	0,45									
$u(f_4, v_8)$	0,55									
$g(v_1)$	0,24	0,27	0,30	0,33	0,36	0,39	0,42	0,45	0,48	0,51
$g(v_2)$	0,16	0,18	0,20	0,22	0,24	0,26	0,28	0,30	0,32	0,34
$g(v_7)$	0,33	0,37	0,41	0,45	0,50	0,54	0,58	0,62	0,66	0,70
$g(v_8)$	0,41	0,46	0,51	0,56	0,62	0,67	0,72	0,77	0,82	0,87
l_1	0,74	0,83	0,92	**1,01**	**1,11**	**1,20**	**1,29**	**1,38**	**1,48**	**1,57**
l_2	0,45	0,50	0,56	0,61	0,67	0,72	0,78	0,84	0,89	0,95
l_5	0,30	0,33	0,37	0,41	0,44	0,48	0,52	0,55	0,59	0,63
l_6	0,33	0,37	0,41	0,45	0,50	0,54	0,58	0,62	0,66	0,70
l_7	0,41	0,46	0,51	0,56	0,62	0,67	0,72	0,77	0,82	0,87
l_{10}	0,41	0,46	0,51	0,56	0,62	0,67	0,72	0,77	0,82	0,87
l_{11}	0,24	0,27	0,30	0,33	0,36	0,39	0,42	0,45	0,48	0,51
l_{14}	0,16	0,18	0,20	0,22	0,24	0,26	0,28	0,30	0,32	0,34
l_{16}	0,33	0,37	0,41	0,45	0,50	0,54	0,58	0,62	0,66	0,70
l_{19}	0,41	0,46	0,51	0,56	0,62	0,67	0,72	0,77	0,82	0,87

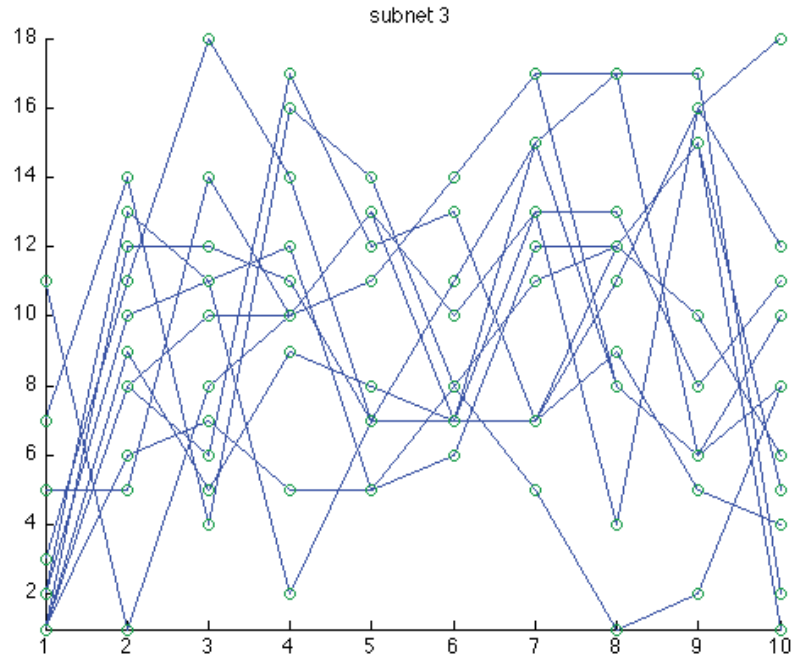
Table 5. Sub-network 1: Optimal Solutions under load balancing control.

4.4 Decomposition evaluation

With the purpose of evaluating the decomposition strategy, in this simulation set we consider randomly generated networks, flows and paths, and use the decomposition algorithm to partition the network in sub-networks. The networks were generated starting from a grid of nodes; in particular, the considered network width is 10 nodes. Each column of the grid can be assigned a number of nodes; in the considered network, the number of nodes per column is [18, 18, 18, 16, 10, 10, 16, 18, 18, 18]. 30 flows were considered, starting from a random node of the first column of the network and directed to a random node of the last column. Similarly, each network path is directed from a node of the first column of the network and directed to a node of the last column Fig. 4 a) shows an example of randomly generated network, whereas Fig. 4 a) shows an example of sub-network. The results were obtained by averaging 20 simulations. The average number of variables of the original problem (i.e., the non-decomposed one) is 1984.8, whereas the decomposition manages to decompose the network in 10.2 sub-network (in the average): each sub-network optimization problem has therefore 194.6 variables, i.e., each sub-network problem is reduced by about one order of magnitude.



a)



b)

Fig. 4. a) example of a network (width=10, height=18), b) one of the sub-networks resulting from the decomposition of the network in Fig. 4 a).

5. Conclusion

In this work we formulate the multipath routing problem as an optimal control problem considering various performance indices. In particular, the scenario includes the load balancing problem already dealt with in a previous work Bruni *et al.*, 2010, as well as the bottleneck minimax control problem, in which the traffic load of the bottleneck (raised to a given power m) is minimized. The mathematical structure of the problem might easily suggest some issues which are evidentiated by the results of Section 4, simply intended to provide a numerical example of more general behaviours. On one side, the load balancing performance index obviously allows to achieve a higher uniformity in the loading of the various links, but it cannot prevent overloading of possible ballast links (apart from *ad hoc* modifications suggested in Bruni *et al.*, 2010).

On the other side, the minimax (bottleneck) approach succeeds in keeping the bottleneck loads (including the ones of the ballast links), as low as possible, with an effort which happens to be more successful the higher the value of m is. This allows accommodating for a higher traffic flow.

Moreover, we stress the fact that the choice of the proper performance index is a matter left to the network manager in charge of the routing control problem, who will have to take into account at the same time the network structure and capacity, as well as the admitted traffic flow and the possible presence of ballast links.

As a final conclusion, we have considered several cost functions for the multipath routing which are suitable for a certain network load situation. Those cost functions can be properly switched during the operations according to the network needs. In that way our approach is strongly oriented with the most innovative vision of the Future Internet perspective (see Delli Priscoli, 2010), in which the core idea is to take consistent and coordinated decisions according to the present contest.

6. References

Bruni, C., Delli Priscoli, F., Koch, G., Marchetti, I. (2009). Resource management in network dynamics: An optimal approach to the admission control problem, Computers & Mathematics with Applications, article in press, available at www.sciencedirect.com, 8 September 2009,doi:10.1016/j.camwa.2009.01.046

Bruni, C., Delli Priscoli, F., Koch, G., Marchetti, I. (2009,b)"Optimal Control of Connection Admission in Telecommunication Networks", European Conrol Conference (ECC) 09, Budapest (Hungary), pp. 2929-2935.

Bruni, C., Delli Priscoli, F., Koch, G., Pietrabissa, A., Pimpinella, L., (2010) "Multipath Routing by Network Decomposition and Traffic Balancing", Proceedings Future Network and Mobile Summit.

Holma H., Toskala A., (2002) WCDMA for UMTS, 2nd Edition.

Warren, A. D., Lasdon, L. S., Suchman, D. F., (1967) Optimization in engineering design, Proc. IEEE, 1885-1897.

Osborne, M. R., Watson, G. A., (1969) An Algorithm for minimax approximation in the non-linear case, Comput. J., 12, pp. 63-68.

Bandler, J. W., Srinivasan, T.V., Charalambous, (1972) Minimax Optimization of networks by Grazor Search, IEEE Trans. Microwave Theory Tech., MTT-20, 596-604.

Bandler, J. W., Charalambous, C. (1972), Practical least pth optimization of networks, IEEE Trans. Microwave Theory Tech., MTT-20, 834-840.

Brayton, R.K., S.W. Director, G.D. Hachtel, and L.Vidigal, (1979), A New Algorithm for Statistical Circuit Design Based on Quasi-Newton Methods and Function Splitting, IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 784-794. Demyanov, V. F., Malozemov, V. N., (1974) Introduction to minimax, John Wiley & Sons.

Cidon, I., Rom R., Shavitt Y., (1999) Analysis of Multipath Routing, IEEE/ACM Transactions ON Networking, Vol. 7, No. 6, pp. 885-896

Banner, R., Orda A., (2007), Multipath Routing Algorithms for Congestion Minimization, IEEE/ACM Transactions on Networking, Vol. 15, No. 2, pp. 413-424

Chen, J., Chan, S.-H. Gary, Li V. O. K., (2004) Multipath Routing for Video Delivery Over Bandwidth-Limited Networks, IEEE Journal on Selected Areas in Communications, Vol. 22, No. 10, pp. 1920-1932

Lin, X., Shroff, N. B., (2006)"Utility Maximization for Communication Networks With Multipath Routing", IEEE Transactions on Automatic Control, Vol. 51, No. 5, pp. 766-781

Güven, T., La, R. J., Shayman, M. A., Bhattacharjee, B., (2008), A Unified Framework for Multipath Routing for Unicast and Multicast Traffic, IEEE/ACM Transactions on Networking, Vol. 16, No. 5, pp. 1038-1051

Jaffe J. M., "Bottlneck Flow Control", IEEE Transactions on Communications, Vol 29, No 7, July 1981, pp. 954-962

Tsai, D., Liau, T. C., Tsai, Wei K., (2006) Least Square Approach to Multipath Maxmin Rate Allocation, 14th IEEE International Conference on Networks, 2006 (ICON '06), Vol. 1, pp. 1-6.

Tsai, Wei K., Kim, Y., (1999) Re-Eximining Maxmin Protocols: A Fundamental Study on Convergence, Complexity, Variations and Performance, 18th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99), Vol. 2, March 1999, pp. 811-818

Delli Priscoli, F., (2010) A Fully Cognitive Approach for Future Internet, Future Internet ISSN 1999-5903 available at www.mdpi.com/journal/futureinternet.

18

Simulation and Optimal Routing of Data Flows Using a Fluid Dynamic Approach

Ciro D'Apice[1], Rosanna Manzo[1] and Benedetto Piccoli[2]
[1]*Department of Electronic and Information Engineering, University of Salerno, Fisciano (SA)*
[2]*Department of Mathematical Sciences, Rutgers University, Camden, New Jersey*
[1]*Italy*
[2]*USA*

1. Introduction

There are various approaches to telecommunication and data networks (see for example Alderson et al. (2007), Baccelli et al. (2006), Baccelli et al. (2001), Kelly et al. (1998), Tanenbaum (1999), Willinger et al. (1998)). A first model for data networks, similar to that used for car traffic, has been proposed in D'Apice et al. (2006), where two algorithms for dynamics at nodes were considered and existence of solutions to Cauchy Problems was proved. Then in D'Apice et al. (2008), following the approach of Garavello et al. (2005) for road networks (see also Coclite et al. (2005); Daganzo (1997); Garavello et al. (2006); Holden et al. (1995); Lighthill et al. (1955); Newell (1980); Richards (1956)), sources and destinations have been introduced, thus taking care of the packets paths inside the network.

In this Chapter we deal with the fluid-dynamic model for data networks together with optimization problems, reporting some results obtained in Cascone et al. (2010); D'Apice et al. (2006; 2008; 2010).

A telecommunication network consists in a finite collection of transmission lines, modelled by closed intervals of $\mathbb{R}$ connected by nodes (routers, hubs, switches, etc.). Taking the Internet network as model, we assume that:

1) Each packet seen as a particle travels on the network with a fixed speed and with assigned final destination;
2) Nodes receive, process and then forward packets which may be lost with a probability increasing with the number of packets to be processed. Each lost packet is sent again.

Since each lost packet is sent again until it reaches next node, looking at macroscopic level, it is assumed that the packets number is conserved. This leads to a conservation law for the packets density ρ on each line:

$$\rho_t + f(\rho)_x = 0. \tag{1}$$

The flux $f(\rho)$ is given by $v(\rho) \cdot \rho$ where v is the average speed of packets among nodes, derived considering the amount of packets that may be lost.

The key point of the model is the loss probability, used to define the flux function. Indeed the choice of a non reasonable loss probability function could invalidate the model. To achieve the goal of the validation of the model assumptions, the loss probability function has been compared with the behaviour of the packet loss derived from known models used in literature to infer network performance and the shape of the velocity and flux functions has been discussed. All the comparisons confirm the validity of the assumptions underlying the fluid-dynamic model (see D'Apice et al. (2010)).

To describe the evolution of networks in which many lines intersect, Riemann Problems (RPs) at junctions were solved in D'Apice et al. (2006) proposing two different routing algorithms:

(RA1) Packets from incoming lines are sent to outgoing ones according to their final destination (without taking into account possible high loads of outgoing lines);

(RA2) Packets are sent to outgoing lines in order to maximize the flux through the node.

One of the drawback of (RA2) is that it does not take into account the global path of packets, therefore leading to possible cycling to bypass congested nodes. These cyclings are avoided if we consider that the packets originated from a source and with an assigned destination have paths inside the network.

Taking this in mind the model was refined in D'Apice et al. (2008). On each transmission line a vector π describing the traffic types, i.e. the percentages of packets going from a source to a destination, has been introduced. Assuming that packets velocity is independent from the source and the destination, the evolution of π follows a semilinear equation

$$\pi_t + v(\rho)\pi_x = 0, \tag{2}$$

hence inside transmission lines the evolution of π is influenced by the average speed of packets.

Different distribution traffic functions describing different routing strategies have been analysed:

- at a junction the traffic started at source s and with d as final destination, coming from the transmission line i, is routed on an assigned line j;
- at a junction the traffic started at source s and with d as final destination, coming from the transmission line i, is routed on every outgoing lines or on some of them.

In particular two ways according to which the traffic at a junction is splitted towards the outgoing lines have been defined. Starting from the distribution traffic function, and using the vector π, the traffic distribution matrix, which describes the percentage of packets from an incoming line that are addressed to an outgoing one, has been assigned. Then, methods to solve RPs according to the routing algorithms (RA1) and (RA2) have been proposed. Optimizations results have been obtained for the model consisting of the conservation law (1). In particular priority parameters and traffic distribution coefficients have been considered as controls and two functionals to measure the efficiency of the network have been defined in Cascone et al. (2010):

1) The velocity of packets travelling through the network.
2) The travel time taken by packets from source to destination.

Due to the nonlinear relation among cost functionals, the optimization of velocity and travel time can give different control parameters.

The analytical treatment of a complex network is very hard due to the high nonlinearity of the dynamics and discontinuities of the I/O maps. For these reasons, a decentralized strategy has been adapted as follows:

Step 1. The optimal controls for asymptotic costs in the case of a single node with constant initial data is computed.

Step 2. For a complex network, the (locally) optimal parameters at every node are used. Thus, the optimal control is determined at each node independently.

The optimization problem for nodes of 2×2 type, i.e. with two entering and two exiting lines, and traffic distribution coefficient α and priority parameter p as control parameters, constant initial data and asymptotic functionals has been completely solved.

Then a test telecommunication network, consisting of 24 nodes, each one of 2×2 type has been studied. Three different choices have been tested for the traffic distribution coefficients and priority parameters: (locally) optimal, static random and dynamic random. The first choice is given by Step 1. By static random parameters, we mean a random choice done at the beginning of the simulation and then kept constant. Finally, dynamic random coefficients are chosen randomly at every instant of time for every node.
The results present some interesting features: the performances of the optimal coefficients are definitely superior with respect to the other two. Then, how the dynamic random choice, which sometimes is equal in performance to the optimal ones, may be not feasible for modelling and robustness reasons has been discussed.

The Chapter is organized as follows. Section 2 reports the model for data networks. Then, in Section 3, we consider possible choices of the traffic distribution functions, and how to compute the traffic distribution matrix from the latter functions and the traffic-type function. We describe two routing algorithms, giving explicit unique solutions to RPs. In Section 4, we discuss the validity of the assumption on the loss probability function, the velocity and flux. The subsequent Section 5 is devoted to the analysis of the optimal control problem introducing the cost functionals. Simulations for three different choices of parameters (optimal, static and dynamic random) in the case of a complex network are presented. The paper ends with conclusions in Section 6.

2. Basic definitions

A telecommunication network is a finite collection of transmission lines connected together by nodes, some of which are sources and destinations. Formally we introduce the following definition:

Definition 1. *A telecommunication network is given by a 7-tuple* $(N, \mathcal{I}, \mathcal{F}, \mathcal{J}, \mathcal{S}, \mathcal{D}, \mathcal{R})$ *where*

Cardinality *N is the cardinality of the network, i.e. the number of lines in the network;*

Lines *$\mathcal{I}$ is the collection of lines, modelled by intervals $I_i = [a_i, b_i] \subseteq \mathbb{R}$, $i = 1, ..., N$;*

Fluxes *$\mathcal{F}$ is the collection of flux functions $f_i : [0, \rho_i^{\max}] \mapsto \mathbb{R}$, $i = 1, ..., N$;*

Nodes *$\mathcal{J}$ is a collection of subsets of $\{\pm 1, ..., \pm N\}$ representing nodes. If $j \in J \in \mathcal{J}$, then the transmission line $I_{|j|}$ is crossing at J as incoming line (i.e. at point b_i) if $j > 0$ and as outgoing line*

(i.e. at point a_i) if $j < 0$. For each junction $J \in \mathcal{J}$, we indicate by $\mathrm{Inc}(J)$ *the set of incoming lines, that are I_i's such that $i \in J$, while by* $\mathrm{Out}(J)$ *the set of outgoing lines, that are I_i's such that $-i \in J$. We assume that each line is incoming for (at most) one node and outgoing for (at most) one node;*

Sources *$\mathcal{S}$ is the subset of $\{1, ..., N\}$ representing lines starting from traffic sources. Thus, $j \in \mathcal{S}$ if and only if j is not outgoing for any node. We assume that $\mathcal{S} \neq \emptyset$;*

Destinations *$\mathcal{D}$ is the subset of $\{1, ..., N\}$ representing lines leading to traffic destinations, Thus, $j \in \mathcal{D}$ if and only if j is not incoming for any node. We assume that $\mathcal{D} \neq \emptyset$;*

Traffic distribution functions *$\mathcal{R}$ is a finite collection of functions (also multivalued) $r_J : \mathrm{Inc}(J) \times \mathcal{S} \times \mathcal{D} \to \mathrm{Out}(J)$. For every J, $r_J(i, s, d)$ indicates the outgoing direction of traffic that started at source s has d as final destination and reached J from the incoming road i.*

2.1 Dynamics on lines

Following D'Apice et al. (2008), we recall the model used to define the dynamics of packet densities along lines. We make the following hypothesis:

(H1) Lines are composed of consecutive processors N_k, which receive and send packets. The packets number at N_k is indicated by $R_k \in [0, R_{max}]$;

(H2) There are two time-scales: Δt_0, the physical travel time of a single packet from node to node (assumed to be independent of the node for simplicity); T, the processing time, during which each processor tries to operate the transmission of a given packet;

(H3) Each processor N_k tries to send all packets R_k at the same time. Packets are lost according to a loss probability function $p : [0, R_{max}] \to [0, 1]$, computed at R_{k+1}, and lost packets are sent again for a time slot of length T;

(H4) The number of packets not transmitted for a whole processing time slot is negligible.

Since the packet transmission velocity on the line is assumed constant, it is possible to compute an average velocity function and thus an average flux function.

Let us focus on two consecutive nodes N_k and N_{k+1}, assume a static situation, i.e. R_k and R_{k+1} are constant. Indicate by δ the distance between the nodes, Δt_{av} the packets average transmission time, $\bar{v} = \frac{\delta}{\Delta t_0}$ the packet velocity without losses and $v = \frac{\delta}{\Delta t_{av}}$ the average packets velocity. Then, we can compute:

$$\Delta t_{av} = \sum_{n=1}^{M} n \Delta t_0 (1 - p(R_{k+1})) p^{n-1}(R_{k+1}),$$

where $M = [T/\Delta t_0]$ (here $[\cdot]$ indicates the floor function) represents the number of attempts of sending a packet and T is the length of a processing time slot. The hypothesis (H4) corresponds to assume $\Delta t_0 << T$ or, equivalently, $M \sim +\infty$. Making the identification, $M = +\infty$, we get:

$$\Delta t_{av} = \frac{\Delta t_0}{1 - p(R_{k+1})},$$

and

$$v = \bar{v}(1 - p(R_{k+1})). \tag{3}$$

Let us call now ρ the averaged density and ρ_{max} its maximum. We can interpret the probability loss function p as a function of ρ and, using (3), determine the corresponding flux function,

given by the averaged density times the average velocity. A possible choice of p is the following:

$$p(\rho) = \begin{cases} 0, & 0 \le \rho \le \sigma, \\ \frac{\rho_{max}(\rho-\sigma)}{\rho(\rho_{max}-\sigma)}, & \sigma \le \rho \le \rho_{\max}, \end{cases} \tag{4}$$

from which

$$v(\rho) = \begin{cases} \bar{v}, & 0 \le \rho \le \sigma, \\ \bar{v}\frac{\sigma(\rho_{max}-\rho)}{\rho(\rho_{max}-\sigma)}, & \sigma \le \rho \le \rho_{\max}, \end{cases} \tag{5}$$

and

$$f(\rho) = \begin{cases} \bar{v}\rho, & 0 \le \rho \le \sigma, \\ \frac{\bar{v}\sigma(\rho_{max}-\rho)}{\rho_{max}-\sigma}, & \sigma \le \rho \le \rho_{\max}. \end{cases} \tag{6}$$

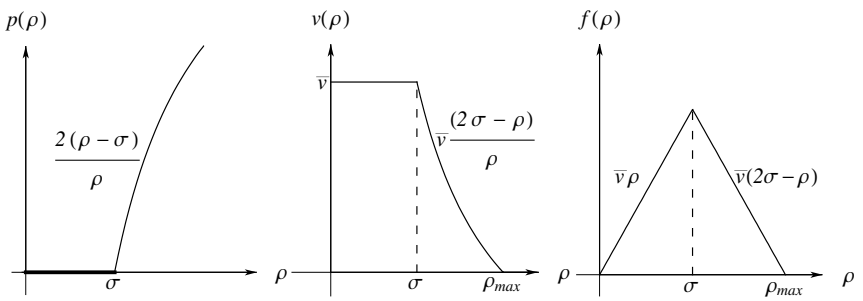


Fig. 1. Loss probability, average velocity and flux behaviours for $\rho_{max} = 1$, $\sigma = \frac{1}{2}$, $\bar{v} = 1$.

To simplify the treatment of the corresponding conservation laws, we will assume the following:

(*F*) Setting $\rho_{max} = 1$, on each line the flux $f : [0,1] \to R$ is concave, $f(0) = f(1) = 0$ and there exists a unique maximum point $\sigma \in]0,1[$.

Notice that the "tent" function

$$f(\rho) = \begin{cases} \rho, & 0 \le \rho \le \frac{1}{2}, \\ 1-\rho, & \frac{1}{2} \le \rho \le 1, \end{cases} \tag{7}$$

and the parabolic flux

$$f(\rho) = \rho(1-\rho), \ \rho \in [0,1], \tag{8}$$

satisfy the assumption (F).

2.2 Dynamics on the network

On each transmission line I_i we consider the evolution equation

$$\partial_t \rho_i + \partial_x f_i(\rho_i) = 0, \tag{9}$$

where we use the assumption (*F*). Therefore, the network load evolution is described by a finite set of functions $\rho_i : [0, +\infty[\times I_i \mapsto [0, \rho_i^{\max}]$.

Moreover, inside each line I_i we define a traffic-type function π_i, which measures the portion of the whole density coming from each source and travelling towards each destination:

Definition 2. *A traffic-type function on a line I_i is a function*

$$\pi_i : [0,\infty[\times[a_i,b_i]\times\mathcal{S}\times\mathcal{D}\mapsto[0,1]$$

such that, for every $t\in[0,\infty[$ and $x\in[a_i,b_i]$

$$\sum_{s\in\mathcal{S},d\in\mathcal{D}}\pi_i(t,x,s,d)=1.$$

In other words, $\pi_i(t,x,s,d)$ specifies the density fraction $\rho_i(t,x)$ that started from source s and is moving towards the final destination d.

Assuming, on the discrete model, that a FIFO policy is used at nodes, it is natural that the averaged velocity, obtained in the limit procedure, is independent from the original sources of packets and their final destinations. In other words, we make the following hypothesis:

(H5) On each line I_i, the average velocity of packets depends only on the value of the density ρ_i and not on the values of the traffic-type function π_i.

As a consequence of hypothesis (H5), we deduce the semilinear equation

$$\partial_t\pi_i(t,x,s,d)+\partial_x\pi_i(t,x,s,d)\cdot v_i(\rho_i(t,x))=0. \tag{10}$$

This equation is coupled with equation (9) on each line I_i. More precisely, equation (10) depends on the solution of (9), while in turn at junctions the values of π_i will determine the traffic distribution on outgoing lines as explained below.

For simplicity and without loss of generality, we assume from now on that the fluxes f_i are all the same and we indicate them with f. Thus, the model for a single transmission line, consists in the system of equations:

$$\begin{cases}\rho_t+f(\rho)_x=0,\\ \pi_t+\pi_x\cdot v(\rho)=0.\end{cases}$$

To treat the evolution at junctions, let us introduce some notations. Fix a junction J with n incoming transmission lines, say $I_1,...,I_n$, and m outgoing transmission lines, say $I_{n+1},...,I_{n+m}$ (junction of $n\times m$ type). The basic ingredient for the solution of Cauchy Problems by Wave Front Tracking method is the solution of Riemann Problems (RPs) (see Bressan (2000), Dafermos (1999), Serre (1999)).

We call RP for a junction the Cauchy Problem corresponding to an initial data $\rho_{1,0},...,\rho_{n+m,0}\in[0,1]$, and $\pi_1^{s,d},...,\pi_{n+m}^{s,d}\in[0,1]$ which are constant on each transmission line.

Definition 3. *A Riemann Solver (RS) for the junction J is a map that associates to Riemann data $\rho_0=(\rho_{1,0},\dots,\rho_{n+m,0})$ and $\Pi_0=(\pi_{1,0},\dots,\pi_{n+m,0})$ at J the vectors $\hat\rho=(\hat\rho_1,\dots,\hat\rho_{n+m})$ and $\hat\Pi=(\hat\pi_1,\dots,\hat\pi_{n+m})$ so that the solution on an incoming transmission line I_i, $i=1,\dots,n$, is given by the wave $(\rho_{i,0},\hat\rho_i)$ and on an outgoing one I_j, $j=n+1,\dots,n+m$, is given by the waves $(\hat\rho_j,\rho_{j,0})$ and $(\hat\pi_j,\pi_{j,0})$. We require the following consistency condition:*

(CC) $RS(RS(\rho_0,\Pi_0))=RS(\rho_0,\Pi_0)$.

Once a RS is defined and the solution of the RP is obtained, we can define admissible solutions at junctions.

3. Riemann Solvers at junctions

Consider a junction J of $n \times m$ type. We denote with $\rho_i(t,x), i = 1,...,n$ and $\rho_j(t,x), j = n+1,...,n+m$ the traffic densities, respectively, on the incoming transmission lines and on the outgoing ones and by $(\rho_{1,0},...,\rho_{n+m,0})$ the initial datum.

Define the maximum flux that can be obtained by a single wave solution on each transmission line as follows:

$$\gamma_i^{\max} = \begin{cases} f(\rho_{i,0}), \text{ if } \rho_{i,0} \in [0,\sigma], \\ f(\sigma), \text{ if } \rho_{i,0} \in]\sigma,1], \end{cases} i = 1,...,n, \tag{11}$$

and

$$\gamma_j^{\max} = \begin{cases} f(\sigma), \text{ if } \rho_{j,0} \in [0,\sigma], \\ f(\rho_{j,0}), \text{ if } \rho_{j,0} \in]\sigma,1], \end{cases} j = n+1,...,n+m. \tag{12}$$

Finally denote with

$$\Omega_i = [0,\gamma_i^{\max}], i = 1,...,n,$$
$$\Omega_j = [0,\gamma_j^{\max}], j = n+1,...,n+m,$$

and with $\hat{\gamma}_{inc} = (f(\hat{\rho}_i),...,f(\hat{\rho}_n)), \hat{\gamma}_{out} = (f(\hat{\rho}_{n+1}),...,f(\hat{\rho}_{n+m}))$ where $\hat{\rho} = (\hat{\rho}_1,...,\hat{\rho}_{n+m})$ is the solution of the RP at the junction.

Now, we discuss some possible choices for the traffic distribution function:

1) $r_J : \mathrm{Inc}(J) \times \mathcal{S} \times \mathcal{D} \to \mathrm{Out}(J)$;
2) $r_J : \mathrm{Inc}(J) \times \mathcal{S} \times \mathcal{D} \hookrightarrow \mathrm{Out}(J)$, i.e. r_J is a multifunction.

If r_J is of type 1), then each packet has a deterministic route, it means that, at the junction J, the traffic that started at source s and has d as final destination, coming from the transmission line I_i, is routed on an assigned line I_j ($r_J(i,s,d) = j$).

Instead if r_J is of type 2), at the junction J, the traffic with source s and destination d coming from a line I_i is routed on every line $I_j \in \mathrm{Out}(J)$ or on some lines $I_j \in \mathrm{Out}(J)$. We can define $r_J(i,s,d)$ in two different ways:

2a) $r_J : \mathrm{Inc}(J) \times \mathcal{S} \times \mathcal{D} \hookrightarrow \mathrm{Out}(J)$,
$r_j(i,s,d) \subseteq \mathrm{Out}(J)$;

2b) $r_J : \mathrm{Inc}(J) \times \mathcal{S} \times \mathcal{D} \to [0,1]^{\mathrm{Out}(J)}$,
$r_J(i,s,d) = (\alpha_J^{i,s,d,n+1},...,\alpha_J^{i,s,d,n+m})$
with $0 \le \alpha_J^{i,s,d,j} \le 1, j \in \{n+1,...,n+m\}, \sum_{j=n+1}^{n+m} \alpha_J^{i,s,d,j} = 1.$

In case 2a) we have to specify in which way the traffic at junction J is splitted towards the outgoing lines.

The definition 2b) means that, at the junction J, the traffic with source s and destination d coming from line I_i is routed on the outgoing line $I_j, j = n+1,...,n+m$ with probability $\alpha_J^{i,s,d,j}$.

Let us analyze how the distribution matrix A is constructed using π and r_J.

Definition 4. *A distribution matrix is a matrix*

$$A \doteq \left\{\alpha_{j,i}\right\}_{j=n+1,...,n+m,i=1,...,n} \in \mathbb{R}^{m\times n}$$

such that

$$0 < \alpha_{j,i} < 1, \quad \sum_{j=n+1}^{n+m} \alpha_{j,i} = 1,$$

for each $i = 1,...,n$ and $j = n+1,...,n+m$, where $\alpha_{j,i}$ is the percentage of packets arriving from the i-th incoming transmission line that take the j-th outgoing transmission line.

In case 1) we can define the matrix A in the following way. Fix a time t and assume that for all $i \in \mathrm{Inc}(J), s \in \mathcal{S}$ and $d \in \mathcal{D}, \pi_i(t,\cdot,s,d)$ admits a limit at the junction J, i.e left limit at b_i. For $i \in \{1,...,n\}, j \in \{n+1,...,n+m\}$, we set

$$\alpha_{j,i} = \sum_{\substack{s\in\mathcal{S},d\in\mathcal{D},\\ r_J(i,s,d)=j}} \pi_i(t,b_i-,s,d).$$

The fluxes $f_i(\rho_i)$ to be consistent with the traffic-type functions must satisfy the following relation:

$$f_j(\rho_j(\cdot,a_j+)) = \sum_{i=1}^{n} \alpha_{j,i} f_i(\rho_i(\cdot,b_i-)),$$

for each $j = n+1,...,n+m$.

Let us analyze how to define the matrix A in the case 2a). We may assign $\varphi(i,s,d) \in r_J(i,s,d)$ and set

$$\alpha_{j,i} = \sum_{\substack{s\in\mathcal{S},d\in\mathcal{D},\\ i:\varphi(i,s,d)=j}} \pi_i(t,b_i-,s,d),$$
$$\alpha_{j,i} = 0, \text{ if } j \notin r_J(i,s,d).$$

However, it is more natural to assign a flexible strategy defining a set of admissible matrices A in the following way

$$\mathcal{A} = \left\{ \begin{array}{c} A : \exists \alpha_J^{i,s,d,j} \in [0,1], \sum_{j=n+1}^{n+m} \alpha_J^{i,s,d,j} = 1, \alpha_J^{i,s,d,j} = 0, \text{ if } j \notin r_J(i,s,d) : \\ \alpha_{j,i} = \sum_{\substack{s\in\mathcal{S},d\in\mathcal{D},\\ j\in r_J(i,s,d)}} \pi_i(t,b_i-,s,d)\alpha_J^{i,s,d,j} \end{array} \right\}.$$

Finally, we treat now the case 2b). In this case the matrix A is unique and is defined by

$$\alpha_{j,i} = \sum_{s\in\mathcal{S},d\in\mathcal{D}} \pi_i(t,b_i-,s,d)\alpha_J^{i,s,d,j}. \tag{13}$$

We describe two different RSs at a junction that represent two different routing algorithms:

(RA1) We assume that

(A) the traffic from incoming transmission lines is distributed on outgoing transmission lines according to fixed coefficients;

(B) respecting (A) the router chooses to send packets in order to maximize fluxes (i.e., the number of packets which are processed).

(RA2) We assume that the number of packets through the junction is maximized both over incoming and outgoing lines.

3.1 Algorithm (RA1)

We have to distinguish case 2a) and 2b).

In case 2a) first we observe that the set $\mathcal{A}$ is convex. The admissible region given by

$$\Omega_{adm} = \{\hat{\gamma} : \hat{\gamma} \in \Omega_1 \times ... \times \Omega_n, \exists A \in \mathcal{A}\ t.c. A\hat{\gamma} \in \Omega_{n+1} \times ... \times \Omega_{n+m}\},$$

is convex at least for the case of junctions of 2×2.

If the region Ω_{adm} is convex than rules (A) and (B) amount to the Linear Programming problem:

$$\max_{\hat{\gamma} \in \Omega_{adm}} (\hat{\gamma}_1 + \hat{\gamma}_2).$$

This problem has clearly a solution, which may not be unique.

Let us consider the case 2b). We need some more notations.

Definition 5. *Let $\tau : [0,1] \to [0,1]$ be the map such that $f(\tau(\rho)) = f(\rho)$ for every $\rho \in [0,1]$ and $\tau(\rho) \neq \rho$ for every $\rho \in [0,1] \backslash \{\sigma\}$.*

We need some assumption on the matrix A (satisfied under generic conditions for $m = n$). Let $\{e_1, ..., e_n\}$ be the canonical basis of $\mathbb{R}^n$ and for every subset $V \subset \mathbb{R}^n$ indicate by $V^\perp$ its orthogonal. Define for every $i = 1, ..., n$, $H_i = \{e_i\}^\perp$, i.e. the coordinate hyperplane orthogonal to e_i and for every $j = n+1, ..., n+m$ let $\alpha_j = \{\alpha_{j1}, ..., \alpha_{jn}\} \in \mathbb{R}^n$ and define $H_j = \{\alpha_j\}^\perp$. Let $\mathcal{K}$ be the set of indices $k = (k_1, ..., k_l), 1 \leq l \leq n-1$, such that $0 \leq k_1 < k_2 < ... < k_l \leq n+m$ and for every $k \in \mathcal{K}$ set $H_k = \bigcap_{h=1}^{l} H_h$. Letting $\mathbf{1} = (1, ..., 1) \in \mathbb{R}^n$, we assume

(C) for every $k \in \mathcal{K}$, $\mathbf{1} \notin H_k^\perp$.

In case 2b) the following result holds

Theorem 6. *(Theorem 3.1 in Coclite et al. (2005) and 3.2 in Garavello et al. (2005)) Let $(N, \mathcal{I}, \mathcal{F}, \mathcal{J}, \mathcal{S}, \mathcal{D}, \mathcal{R})$ be an admissible network and J a junction of $n \times m$ type. Assume that the flux $f : [0,1] \to \mathbb{R}$ satisfies (F) and the matrix A satisfies condition (C). For every $\rho_{1,0}, ..., \rho_{n+m,0} \in [0,1]$, and for every $\pi_1^{s,d}, ...\pi_{n+m}^{s,d} \in [0,1]$, there exist densities $\hat{\rho}_1, ..., \hat{\rho}_{n+m}$ and a unique admissible centered weak solution, $\rho = (\rho_1, ..., \rho_{n+m})$ at J such that*

$$\rho_1(0, \cdot) \equiv \rho_{1,0}, ..., \rho_{n+m}(0, \cdot) \equiv \rho_{n+m,0},$$
$$\pi^1(0, \cdot s, d) = \pi_1^{s,d}, ..., \pi^{n+m}(0, \cdot, s, d) = \pi_{n+m}^{s,d}, (s \in \mathcal{S}, d \in \mathcal{D}).$$

We have

$$\hat{\rho}_i \in \begin{cases} \{\rho_{i,0}\} \cup \left]\tau(\rho_{i,0}), 1\right], & \text{if } 0 \leq \rho_{i,0} \leq \sigma, \\ [\sigma, 1], & \text{if } \sigma \leq \rho_{i,0} \leq 1, \end{cases} \quad i = 1, ..., n, \tag{14}$$

$$\hat{\rho}_j \in \begin{cases} [0,\sigma], & \text{if } 0 \le \rho_{j,0} \le \sigma, \\ \{\rho_{j,0}\} \cup \left[0, \tau(\rho_{j,0})\right[, & \text{if } \sigma \le \rho_{j,0} \le 1, \end{cases} \quad j = n+1, ..., n+m, \tag{15}$$

and on each incoming line I_i, $i = 1, ..., n$, the solution consists of the single wave $(\rho_{i,0}, \hat{\rho}_i)$, while on each outgoing line I_j, $j = n+1, ..., n+m$, the solution consists of the single wave $(\hat{\rho}_j, \rho_{j,0})$. Moreover $\hat{\pi}_i(t, \cdot, s, d) = \pi_i^{s,d}$ for every $t \ge 0, i \in \{1, ..., n\}, s \in \mathcal{S}, d \in \mathcal{D}$ and

$$\hat{\pi}_j(t, a_j+, s, d) = \frac{\sum_{i=1}^{n} \alpha_J^{i,s,d,j} \pi_i^{s,d}(t, b_i-, s, d) f(\hat{\rho}_i)}{f(\hat{\rho}_j)}$$

for every $t \ge 0, j \in \{n+1, ..., n+m\}, s \in \mathcal{S}, d \in \mathcal{D}$.

3.2 Algorithm (RA2)

To solve RPs according to (RA2) we need some additional parameters called priority and traffic distribution parameters. For simplicity of exposition, consider, junction J of 2×2 type. In this case we have only one priority parameter $q \in]0,1[$ and one traffic distribution parameter $\alpha \in]0,1[$. We denote with $(\rho_{1,0}, \rho_{2,0}, \rho_{3,0}, \rho_{4,0})$ and $(\pi_{1,0}^{s,d}, \pi_{2,0}^{s,d}, \pi_{3,0}^{s,d}, \pi_{4,0}^{s,d})$ the initial data.

In order to maximize the number of packets through the junction over incoming and outgoing lines we define

$$\Gamma = \min\left\{\Gamma_{in}^{\max}, \Gamma_{out}^{\max}\right\},$$

where $\Gamma_{in}^{\max} = \gamma_1^{\max} + \gamma_2^{\max}$ and $\Gamma_{out}^{\max} = \gamma_3^{\max} + \gamma_4^{\max}$. Thus we want to have Γ as flux through the junction.

One easily see that to solve the RP, it is enough to determine the fluxes $\hat{\gamma}_i = f(\hat{\rho}_i), i = 1, 2$. In fact, to have simple waves with the appropriate velocities, i.e. negative on incoming lines and positive on outgoing ones, we get the constraints (14), (15). Observe that we compute $\hat{\gamma}_i = f(\hat{\rho}_i), i = 1, 2$ without taking into account the type of traffic distribution function.

We have to distinguish two cases:

I $\Gamma_{in}^{\max} = \Gamma$,
II $\Gamma_{in}^{\max} > \Gamma$.

In the first case we set $\hat{\gamma}_i = \gamma_i^{\max}, i = 1, 2$.
Let us analyze the second case in which we use the priority parameter q. Not all packets can enter the junction, so let C be the amount of packets that can go through: qC packets come from first incoming line and $(1-q)C$ packets from the second. In the space (γ_1, γ_2), define the following lines:

$$r_q : \gamma_2 = \frac{1-q}{q}\gamma_1, \quad r_\Gamma : \gamma_1 + \gamma_2 = \Gamma,$$

and P the point of intersection of r_q and r_Γ. Recall that the final fluxes should belong to the region:

$$\Omega_{in} = \left\{(\gamma_1, \gamma_2) : 0 \le \gamma_i \le \gamma_i^{\max}, i = 1, 2\right\}.$$

We distinguish two cases:

a) P belongs to Ω_{in},

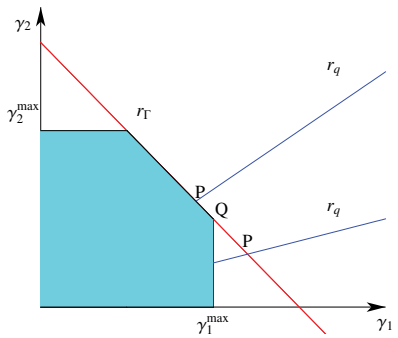


Fig. 2. P belongs to Ω_{in} and P is outside Ω_{in}.

b) P is outside Ω_{in}.

In the first case we set $(\hat{\gamma}_1, \hat{\gamma}_2) = P$, while in the second case we set $(\hat{\gamma}_1, \hat{\gamma}_2) = Q$, with $Q = proj_{\Omega_{in} \cap r_\Gamma}(P)$ where $proj$ is the usual projection on a convex set, see Figure 2.

As for the algorithm (RA1) $\hat{\pi}_i^{s,d} = \pi_{i,0}^{s,d}, i = 1, 2$.

Let us now determine $\hat{\gamma}_j, j = 3, 4$. We have to distinguish again two cases :

I $\Gamma_{out}^{\max} = \Gamma$,
II $\Gamma_{out}^{\max} > \Gamma$.

In the first case $\hat{\gamma}_j = \gamma_j^{\max}, j = 3, 4$. Let us determine $\hat{\gamma}_j$ in the second case, using the traffic distribution parameter α. Since not all packets can go on the outgoing transmission lines, we let C be the amount that goes through. Then αC packets go on the outgoing line I_3 and $(1 - \alpha)C$ on the outgoing line I_4. Consider the space (γ_3, γ_4) and define the following lines:

$$r_\alpha : \gamma_4 = \frac{1 - \alpha}{\alpha} \gamma_3,$$

$$r_\Gamma : \gamma_3 + \gamma_4 = \Gamma.$$

We have to distinguish case 2a) and 2b) for the traffic distribution function.

3.2.1 Case 2a)

Let us introduce the connected set

$$\mathcal{G} = \left\{ A \hat{\gamma}_{inc}^T : A \in \mathcal{A} \right\},$$

and G_1 and G_2 its endpoints. Since in case 2a) we have an infinite number of matrices A, each of one determines a line r_α, we choose the most "natural" line r_α , i.e. the one nearest to the statistic line determined by measurements on the network.

Recall that the final fluxes should belong to the region:

$$\Omega_{out} = \left\{ (\gamma_3, \gamma_4) : 0 \leq \gamma_j \leq \gamma_j^{\max}, j = 3, 4 \right\}.$$

Define $P = r_\alpha \cap r_\Gamma, R = (\Gamma - \gamma_4^{\max}, \gamma_4^{\max}), Q = (\gamma_3^{\max}, \Gamma - \gamma_3^{\max})$. We distinguish 3 cases:

a) $\mathcal{G} \cap \Omega_{out} \cap r_\Gamma \neq \varnothing$,
b) $\mathcal{G} \cap \Omega_{out} \cap r_\Gamma = \varnothing$ and $\gamma_3(G_1) < \gamma_3(R)$,
c) $\mathcal{G} \cap \Omega_{out} \cap r_\Gamma = \varnothing$ and $\gamma_3(G_1) > \gamma_3^{\max}$.

If the set $\mathcal{G}$ has a priority over the line r_Γ we set $(\hat{\gamma}_3, \hat{\gamma}_4)$ in the following way. In case a) we define $(\hat{\gamma}_3, \hat{\gamma}_4) = proj_{\mathcal{G} \cap \Omega_{out} \cap r_\Gamma}(P)$, in case b) $(\hat{\gamma}_3, \hat{\gamma}_4) = R$, and finally in case c) $(\hat{\gamma}_3, \hat{\gamma}_4) = Q$.

Otherwise, if r_Γ has a priority over $\mathcal{G}$ we set $(\hat{\gamma}_3, \hat{\gamma}_4) = \min_{\gamma \in \Omega_{out}} \mathcal{F}(\gamma, r_\alpha, \mathcal{G})$ where $\mathcal{F}$ is a convex functional which depends on γ, r_α and on the set $\mathcal{G}$ of the routing standards.

The vector $\hat{\pi}_i^{s,d}, j = 3, 4$ are computed in the same way as for the algorithm (RA1).

3.2.2 Case 2b)

In case 2b) we have a unique matrix A. The fluxes on outgoing lines are computed as in the case without sources and destinations.

We distinguish two cases:

a) P belongs to Ω,
b) P is outside Ω.

In the first case we set $(\hat{\gamma}_3, \hat{\gamma}_4) = P$, while in the second case we set $(\hat{\gamma}_3, \hat{\gamma}_4) = Q$, where $Q = proj_{\Omega_{adm}}(P)$. Again, we can extend to the case of m outgoing lines.

Finally we define $\hat{\pi}_i^{s,d}, j = 3, 4$ as in the case 2a):

$$\hat{\pi}_j(t, a_j+, s, d) = \frac{\sum_{i=1}^{n} \alpha_J^{i,s,d,j} \pi_i^{s,d}(t, b_i-, s, d) f(\hat{\rho}_i)}{f(\hat{\rho}_j)}$$

for every $t \geq 0, j \in \{n+1, ..., n+m\}, s \in \mathcal{S}, d \in \mathcal{D}$.

Once solutions to RPs are given, one can use a Wave Front Tracking algorithm to construct a sequence of approximate solutions.

4. Model assumptions

The aim of this section is to verify that the assumptions underlying the data networks fluid-dynamic model (shortly FD model) are correct. Here we focus on the fixed-point models to describe TCP, and considering various set-ups with TCP traffic in a single bottleneck topology, we investigate queueing models for estimating packet loss rate. In what follows we suppose $\rho_{max} = 1$ and $\sigma = \frac{1}{2}$.

4.1 Loss probability function

It is reasonable to assume that the loss probability function p is null for some interval, which is a right neighborhood of zero. This means that at low densities no packet is lost. Then p should be increasing, reaching the value 1 at the maximal density, the situation of complete stuck. With the above assumptions the loss probability function in (4) can be written as:

$$p(\rho) = \begin{cases} 0, & 0 \leq \rho \leq 1/2, \\ \frac{2\rho - 1}{\rho}, & 1/2 \leq \rho \leq 1. \end{cases} \tag{16}$$

We analyze some models used in literature to evaluate the packets loss rate with the aim to compare its behaviour with the function depicted in Figure 1.

4.1.1 The proportional-excess model

Let us consider the transmission of two consecutive routers. The node that transmits packets is called *sender*, while the receiving one is said *receiver*. Among the nodes, there is a link or channel, with limited capacity. Assume that the sender and the receiver are synchronized each other, i.e. the receiver is able to process in real time all packets, sent by the sender. In few words, no packets are lost. The packets loss can occur only on the link, due to its finite capacity. Under the zero buffer hypotheses the loss rate is defined as the proportional excess of offered traffic over the available capacity. If R is the sender bit rate and C is the link capacity, we have a loss if $R > C$. The model is said *proportional-excess* or briefly P/E and suppose deterministic arrivals. The packets bit rate is:

$$p = \begin{cases} 0, & R < C, \\ \frac{R-C}{C}, & R > C. \end{cases} \tag{17}$$

In Figure 3, loss probability for P/E model (continuous curve) and FD model (dashed curve) are shown, assuming $C = \sigma = 1/2$. For values $C < \rho < 2C$, the FD model overestimates the loss probability.

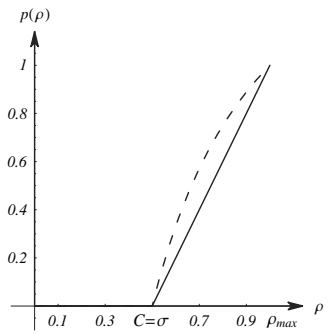


Fig. 3. Loss probabilities. Dashed line: FD model. Continuous line: P/E model.

Observe that the P/E model is not realistic. In fact, the sender and the receiver are never synchronized each other and whatever transmission protocol is used by the transport layer, the receiver has a finite length buffer, where the packets wait to be processed and eventually sent to the next node. Thus queueing models are needed, to infer about network performance.

4.1.2 Models with finite capacity

Queueing models are good at predicting loss in a network with many independent users, probably using different applications. Consider the traffic from TCP sources that send packets through a bottleneck link. The traffic is aggregated and used as an arrival process for the link. The arrival process, being the aggregation of independent sources, is approximated as a Poisson process, and the aggregated throughput is used as the rate of the Poisson process (see Wierman et al. (2003)). These considerations justify the assumption that the times between the packets arrivals are exponentially distributed. Depending on the hypothesis on the length of

packets arriving to the queue the data transmission can be modelled with different queueing models, as $M/D/1/B$ and $M/M/1/B$, characterized by deterministic and exponentially distributed lengths, respectively, and a buffer with capacity $B-1$. From the queue length distribution, known in closed formulas or iteratively in the finite buffer case, expected time in queue and in the system, as well as packet loss rate can be derived. In what follows we denote the arrival intensity by λ, the service intensity by μ and define the load as $\rho = \lambda/\mu$.

4.1.2.1 Fixed packets dimension

In a scenario where all senders use the same data packets size, the queueing model $M/D/1/B$ is the most natural choice. The probability that the buffer is full gives the loss rate:

$$p(\rho) = \frac{1 + (\rho - 1)\,\alpha_B(\rho)}{1 + \rho\alpha_B(\rho)}, \tag{18}$$

where

$$\alpha_B(\rho) = \sum_{k=0}^{B-2} \frac{e^{\rho(B-k-1)}\,(-1)^k\,(B-k-1)^k\,\rho^k}{k!},\ B \geq 2.$$

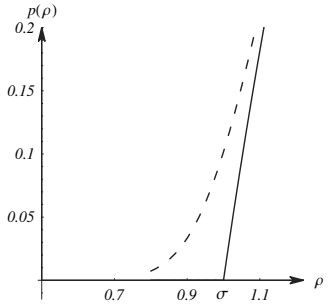


Fig. 4. Loss rates. Dashed line: $M/D/1/B$ model. Continuous line: FD model.

Figure 4 shows a comparison among the loss rate (16) and (18), assuming $B = 10$. However, an $M/D/1/B$ queue predicts a lower loss rate and higher throughput than is seen in the true network. This is due to fact that in real routers packet sizes are not always fixed to the maximum segment size, therefore packet sizes are more variable than a deterministic distribution.

4.1.2.2 Exponentially distributed packets size

Assume the packet size is exponentially distributed. This assumption is true if we consider the total amount of traffic as the superposition of traffic fluxes, coming from different TCP sources, each configured to use its own packet size. The $M/M/1/B$ queue is a good approximation of the simulated bottleneck link shared among TCP sources under any traffic load (Wierman et al. (2003)). The loss rate for the $M/M/1/B$ queueing model is:

$$p(\rho) = \frac{\rho^B\,(1-\rho)}{1-\rho^{B+1}}. \tag{19}$$

In Figure 5, left, the loss bit rate for different values of the buffer ($B = 10, 20, 30$) is reported. Notice that, increasing the B values, dashed lines tend to the continuous one.

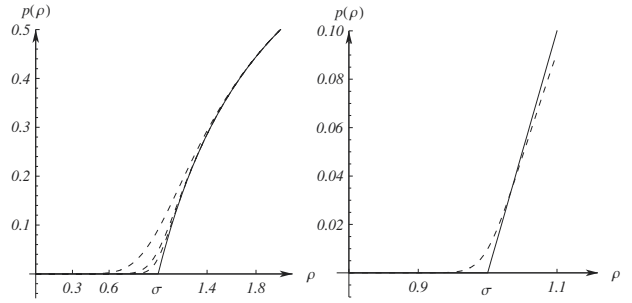


Fig. 5. Left: Loss bit rate for different values of the buffer. Right: Loss probability function. Dashed lines: $M/M/1/B$. Continuous line: P/E model.

In fact, the loss probability of the FD model represents for $\sigma = 1$ (up to a scale factor equal to 2) a limit case of (19):

$$\lim_{B\to\infty} \frac{\rho^B (1-\rho)}{1-\rho^{B+1}} = \begin{cases} 0, & 0 < \rho \leq 1, \\ \frac{\rho-1}{\rho}, & \rho > 1. \end{cases}$$

The loss probability for the queueing model (dashed line) and the P/E one (continuous line) is shown in Figure 5, right. The two curves almost match for small bit rate values, i.e. in the load range $0.9\sigma < \rho < 1.1\sigma$. For greater loads values, the P/E model overestimates the loss probability.

Theoretical and simulative studies pointed out that $M/D/1/B$ and $M/M/1/B$ queueing models give good prediction of the loss rate in network with many independent users performing short file transfers (shorts FTP). In literature other queueing models have been considered to describe different scenarios, as bach arrivals. For a comparison among different models see Figure 6, where the packet loss rate for $M/D/1/B$, $M/M/1/B$, $M^2/M/1/B$, $M^5/M/1/B$ and the P/E models are reported for the case $B = 100$ and loads in the interval $0.8 < \rho < 1.1$. Observe that $M^r/M/1/B$ denotes a queue with Poisson batch arrivals of size r and describes the fact that TCP traffic is likely to be quite bursty due to synchronized loss events that are experienced by multiple users.

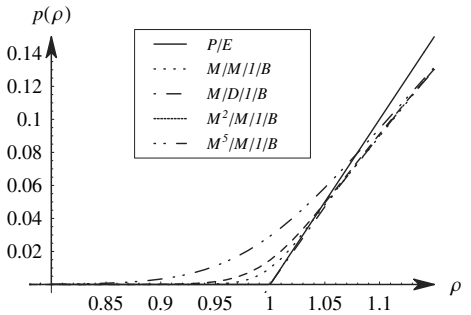


Fig. 6. Comparison of different queueing models.

Significant difference are restricted to the range $0.9\sigma < \rho < 1.1\sigma$. As the load increases above 1.1 the loss estimates become very close in the different queueing models. Any of these models

predict the loss rate equally well. However, under low loss environments, the best queueing model depends on the type of transfers by TCP sources, i.e. persistent or transient. It is shown in Olsen (2003) that $M/D/1/B$ queues estimations of the loss rate can be used for transient sources. However, for sources with a slightly longer on and off periods, $M/M/1/B$ queues best predict the loss rate, and for (homogeneous) persistent sources, $M^r/M/1/B$ queues give better performance inferences, due to the traffic burstiness stemming from the TCP slow-start and source synchronization effect. Even if some models are more appropriate in situations of low load, others when the load is heavy, Figure 6 shows that the assumption on the loss probability function of the FD model is valid.

4.2 Velocity

The loss probability, influencing the average transmission time, has effects on the average velocity of packets:

$$v(\rho) = \bar{v}\,(1 - p(\rho))\,.$$

The behaviour of the average velocity in the FD model

$$v\,(\rho) = \begin{cases} \bar{v}, & 0 \leq \rho \leq /2, \\ \bar{v}\frac{1-\rho}{\rho}, & 1/2 \leq \rho \leq 1, \end{cases} \tag{20}$$

is depicted in Figure 1. Notice that the velocity is constant if the system is free (no losses). Over the threshold, losses occur, and the average travelling time increasing reduces the velocity. The average packet velocity for the P/E model and the $M/M/1/B$ model is plotted in Figure 7. Such two curves fit the curve of the FD model, confirming the goodness of its assumptions.

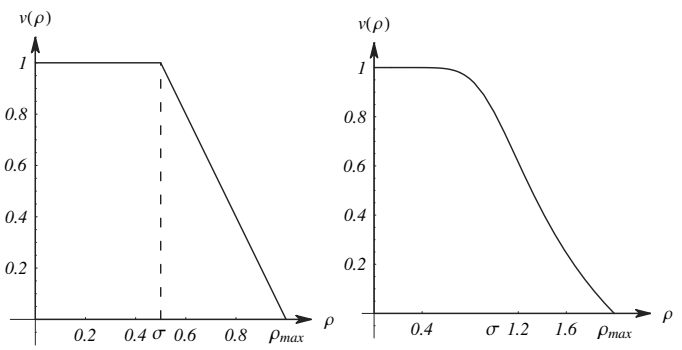


Fig. 7. Average velocity. Left: P/E model. Right: $M/M/1/B$ model.

4.3 Flux

Once the velocity function is known, the flux is given by:

$$f(\rho) = v(\rho)\rho.$$

In case of the FD model

$$f\,(\rho) = \begin{cases} \bar{v}\rho, & 0 \leq \rho \leq 1/2, \\ \bar{v}(1-\rho), & 1/2 \leq \rho \leq 1, \end{cases} \tag{21}$$

see Figure 1. For the P/E model, we get

$$f(\rho) = \begin{cases} \rho\bar{v}, & 0 \leqslant \rho \leqslant \sigma, \\ \frac{(2\sigma-\rho)\bar{v}\rho}{\sigma}, & \sigma \leqslant \rho \leqslant \rho_{\max}. \end{cases} \tag{22}$$

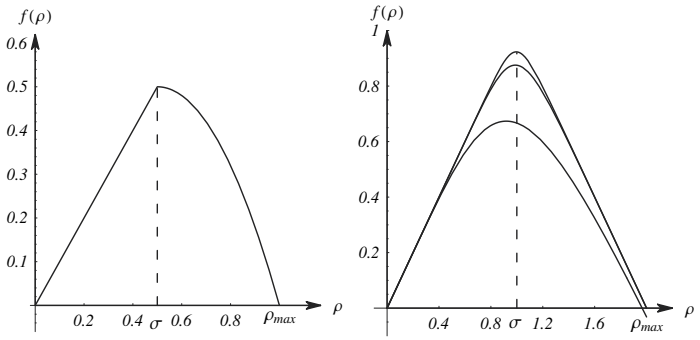


Fig. 8. Flux. Left: P/E model. Right: $M/M/1/B$ (for $B = 5$, $B = 15$, $B = 25$).

The flux in the P/E model and $M/M/1/B$ model are depicted in Figure 8. Note the effects of a finite buffer on the maximal value of the flux. If B tends to infinity, the flux best approximates the FD model flux. For small B values, the maximal flux decreases and the load value in which the maximum is attained is shifted on the right due to the fact that packets are lost for load values smaller than the threshold.

5. Optimal control problems for telecommunication networks

Now we state optimal control problems on the network.
We have a network $(\mathcal{I}, \mathcal{J})$, with nodes of at most 2×2 type, and an initial data $\rho_0 = (\rho_{i,0})_{i=1,..,N}$. The evolution is determined by equation (9) on each line I_i and by Riemann Solvers RS_J, depending on priority and traffic distribution parameters, q and α, respectively. For the definition of RS_J see the case when the traffic distribution function is of type 2b).

We now consider α and q as controls. To measure the efficiency of the network, it is natural to consider two quantities:

1) The average velocity at which packets travel through the network.
2) The average time taken by packets from source to destination.

Clearly, to optimize 1) and 2) is the same if we refer to a single packet, but the averaged values may be very different (since there is a nonlinear relation among the two quantities). As the model consider macroscopic quantities, we can estimate the averages integrating over time and space the average velocity and the reciprocal of average velocity, respectively. We thus define the following:

$$J_1(t) = \sum_i \int_{I_i} v(\rho_i(t,x))\, dx,$$

$$J_2(t) = \sum_i \int_{I_i} \frac{1}{v(\rho_i(t,x))}\, dx,$$

and, to obtain finite values, we assume that the optimization horizon is given by $[0, T]$ for some $T > 0$.
Notice that this corresponds to the following operation:

- average in time and then w.r.t packets, to compute the probability loss function;
- average in space, to pass to the limit and get model (9);
- integrate in space and time to get the final value.

The value of such functionals depends on the order in which averages and integrations are taken.

Summarizing, we get the following optimal control problems:

Data. Network $(\mathcal{I}, \mathcal{J})$; initial data $\bar{\rho} = (\bar{\rho}_i)_{i=1,\dots,N}$; optimization horizon $[0, T]$, $T > 0$.

Dynamics. Equation (9) on each line $I \in \mathcal{I}$ and Riemann Solver RS_J for each $J \in \mathcal{J}$, depending on controls α and q.

Control Variables. Traffic distribution parameter $t \mapsto \alpha_J(t)$ and priority parameter $t \mapsto q_J(t)$, i.e. two controls for every node $J \in \mathcal{J}$.

Control Space. $\{(\alpha_J, q_J) : J \in \mathcal{J},\ \alpha_J, q_J \in L^\infty([0, T], [0, 1])\}$.

Cost functions. Integrated functionals:

$$\max \int_0^T J_1(t)\, dt, \qquad \min \int_0^T J_2(t)\, dt.$$

Definition 7. *We call* $(\mathbf{P}_i)$ *the optimal control problem referred to the functional* J_i*:*

$$(\mathbf{P_1}) \quad \max_{(\alpha,q)} J_1, \text{ subject to (9).}$$

$$(\mathbf{P_2}) \quad \min_{(\alpha,q)} J_2, \text{ subject to (9).}$$

The direct solution of problems $(\mathbf{P}_i)$ corresponds to a centralized approach. We propose the alternative approach of decentralized algorithm more precisely:

Step 1 For every node J and Riemann Solver RS_J, solve the simplified optimal control problem:

$$\max\ (or \min)\ J_i(T),$$

for T sufficiently big, on the network formed only by J with constant initial data, taking approximate solutions when there is lack of existence.

Step 2 Apply the obtained optimal control at every time t in the optimization horizon and at every node J, taking the value at J on each line as initial data.

Notice that, for T sufficiently big, we can assume that the datum is constant on each line: this strongly simplifies the approach.

We consider a single node J with incoming lines, labelled by 1 and 2, and with outgoing lines, labelled by 3 and 4.
Since $\widehat{\rho} = \widehat{\gamma}$, $0 \le \widehat{\rho} \le \frac{1}{2}$, and $\widehat{\rho} = 1 - \widehat{\gamma}$, $\frac{1}{2} \le \widehat{\rho} \le 1$, we have that $v\left(\widehat{\rho}_\varphi\right) = H\left(-s_\varphi\right) +$

$\frac{1-\widehat{\rho}_\varphi}{\widehat{\rho}_\varphi} H\left(s_\varphi\right), \varphi = 1,2, v\left(\widehat{\rho}_\psi\right) = H\left(-s_\psi\right) + \frac{1-\widehat{\rho}_\psi}{\widehat{\rho}_\psi} H\left(s_\psi\right), \psi = 3,4$, where $H(x)$ is the Heavyside function and s_φ and s_ψ are determined by the solution to the RP at J:

$$s_\varphi = \begin{cases} -1, \text{ if } \rho_{\varphi,0} \leq \frac{1}{2} \text{ and } \Gamma = \Gamma_{in}, \text{ or } \rho_{\varphi,0} \leq \frac{1}{2}, q_\varphi \Gamma = \gamma_\varphi^{\max} \text{ and } \Gamma = \Gamma_{out}, \\ +1 \text{ if } \rho_{\varphi,0} > \frac{1}{2}, \text{ or } \rho_{\varphi,0} \leq \frac{1}{2}, q_\varphi \Gamma < \gamma_\varphi^{\max} \text{ and } \Gamma = \Gamma_{out}, \end{cases} \varphi = 1,2,$$

$$s_\psi = \begin{cases} -1, \text{ if } \rho_{\psi,0} < \frac{1}{2}, \text{ or } \rho_{\varphi,0} \geq \frac{1}{2}, \alpha_\psi \Gamma < \gamma_\psi^{\max} \text{ and } \Gamma = \Gamma_{in}, \\ +1 \text{ if } \rho_{\psi,0} \geq \frac{1}{2} \text{ and } \Gamma = \Gamma_{out}, \text{ or } \rho_{\psi,0} \geq \frac{1}{2}, \alpha_\psi \Gamma = \gamma_\psi^{\max} \text{ and } \Gamma = \Gamma_{in}. \end{cases} \psi = 3,4,$$

with:

$$q_\varphi = \begin{cases} q, & \text{if } \varphi = 1, \\ 1-q, & \text{if } \varphi = 2, \end{cases} \quad \alpha_\psi = \begin{cases} \alpha, & \text{if } \psi = 3, \\ 1-\alpha, & \text{if } \psi = 4. \end{cases}$$

Then, for T sufficiently big,

$$J_1(T) = 2\left[v\left(\widehat{\rho}_1\right) + v\left(\widehat{\rho}_2\right) + v\left(\widehat{\rho}_3\right) + v\left(\widehat{\rho}_4\right)\right]; \tag{23}$$

$$J_2(T) = t\left(\widehat{\rho}_1\right) + t\left(\widehat{\rho}_2\right) + t\left(\widehat{\rho}_3\right) + t\left(\widehat{\rho}_4\right), \tag{24}$$

with

$$t\left(\widehat{\rho}_x\right) = \frac{\widehat{\rho}_x}{H\left(s_x\right) + \widehat{\rho}_x\left[H\left(-s_x\right) - H\left(s_x\right)\right]}.$$

We want to maximize the cost $J_1(T)$ and to minimize the cost $J_2(T)$ with respect to the parameters α and q. In Marigo (2006) and Cascone et al. (2007), you can find a similar approach for telecommunication networks and road networks, respectively, modelled with flux function (8). Let

$$\beta^- = \frac{\Gamma - \gamma_3^{\max}}{\gamma_3^{\max}}, \ \beta^+ = \frac{\gamma_4^{\max}}{\Gamma - \gamma_4^{\max}},$$
$$p^- = \frac{\Gamma - \gamma_1^{\max}}{\gamma_1^{\max}}, \ p^+ = \frac{\gamma_2^{\max}}{\Gamma - \gamma_2^{\max}}.$$

Theorem 8. *Consider a junction J of 2×2 type. If $\Gamma = \Gamma_{in} = \Gamma_{out}$ and T is sufficiently big, the cost functionals $J_1(T)$ and $J_2(T)$ depend neither on α nor q. If $\Gamma = \Gamma_{in}$, the cost functionals $J_1(T)$ and $J_2(T)$ depend only on α. The optimal values for $J_1(T)$ are the following:*

(i) *if $s_3 = s_4 = +1$, and $\beta^- \leq 1 \leq \beta^+, \beta^-\beta^+ > 1$, or $1 \leq \beta^- \leq \beta^+, \alpha \in \left[0, \frac{1}{1+\beta^+}\right]$;*

(ii) *if $s_3 = s_4 = +1$, and $\beta^- \leq 1 \leq \beta^+, \beta^-\beta^+ = 1, \alpha \in \left[0, \frac{1}{1+\beta^+}\right[\cup \left]\frac{1}{1+\beta^-}, 1\right]$;*

(iii) *if $s_3 = s_4 = +1$, and $\beta^- \leq 1 \leq \beta^+, \beta^-\beta^+ < 1$, or $\beta^- \leq \beta^+ \leq 1, \alpha \in \left]\frac{1}{1+\beta^-}, 1\right]$;*

(iv) *if $s_3 = -s_4 = -1, \alpha \in \left[0, \frac{1}{1+\beta^+}\right[$ in the cases: $\beta^- \leq 1 \leq \beta^+, 1 \leq \beta^- \leq \beta^+$, or $\beta^- \leq \beta^+ \leq 1$;*

(v) *if $s_3 = -s_4 = +1, \alpha \in \left]\frac{1}{1+\beta^-}, 1\right]$ in the cases: $\beta^- \leq 1 \leq \beta^+, 1 \leq \beta^- \leq \beta^+$, or $\beta^- \leq \beta^+ \leq 1$.*

If $\Gamma = \Gamma_{in}$, the optimal values for $J_2(T)$ are the following:

(i) *if $s_3 = s_4 = +1$ or $s_c = -s_d = -1$, and $\beta^- \leq 1 \leq \beta^+, \alpha = \frac{1}{2}$;*

(ii) *if $s_3 = s_4 = +1$, and $\beta^- \leq \beta^+ \leq 1, \alpha \in \left[0, \frac{1}{1+\beta^+}\right]$;*

(iii) *if $s_3 = s_4 = +1$, and $1 \leq \beta^- \leq \beta^+, \alpha \in \left[\frac{1}{1+\beta^-}, 1\right]$;*

(iv) *if $s_3 = -s_4 = -1$, and $1 \leq \beta^- \leq \beta^+$, or $\beta^- \leq \beta^+ \leq 1, \alpha \in \left[0, \frac{1}{1+\beta^+}\right[$;*

(v) *if* $s_3 = -s_4 = +1$, *and* $\beta^- \leq 1 \leq \beta^+$, *or* $1 \leq \beta^- \leq \beta^+$, *or* $\beta^- \leq \beta^+ \leq 1$, $\alpha \in \left]\frac{1}{1+\beta^-}, 1\right]$.

If $\Gamma = \Gamma_{out}$, *the cost functionals* $J_1(T)$ *and* $J_2(T)$ *depend only on* q. *The optimal values for* $J_1(T)$ *and* $J_2(T)$ *are the same for* α *when* $\Gamma = \Gamma_{in}$, *if we substitute* α *with* q, β^- *with* p^-, *and* β^+ *with* p^+.

5.1 A case study

In what follows, we report the simulation results of a test telecommunication network, that consists of nodes of 2×2 type. The network, represented in Figure 9, is characterized by:

- 24 nodes;
- 12 incoming lines: $1, 2, 5, 8, 9, 16, 19, 20, 31, 32, 45, 46$;
- 12 outgoing lines: $6, 17, 29, 43, 48, 50, 52, 54, 56, 58, 59, 60$;
- 36 inner lines: $3, 4, 7, 10, 11, 12, 13, 14, 15, 18, 21, 22, 23, 24, 25, 26, 27, 28, 30, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 47, 49, 51, 53, 55, 57$.

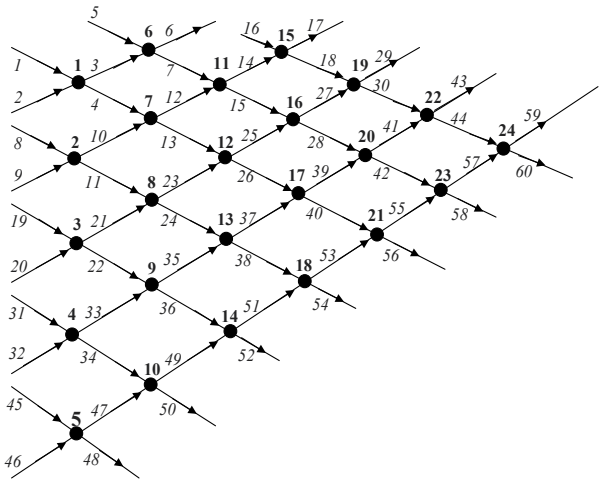


Fig. 9. Network with 24 nodes.

We distinguish three case studies, that can be called, case A, B, and C. In Table 1, we report the initial conditions $\rho_{i,0}$ and the boundary data (if necessary) $\rho_{bi,0}$ for case A.

As for case B, instead, we consider the same initial conditions of case A, but boundary data equal to 0.75.

Table 2 contains initial and boundary conditions for case C. An initial condition of 0.75 is assumed for the inner lines of the network, that are not present in Table 2.

As in Bretti et al. (2006), we consider approximations obtained by the numerical method of Godunov (Godunov (1959)), with space step $\Delta x = 0.0125$ and time step determined by the CFL condition (Godlewsky et al. (1996)). The telecommunication network is simulated in a time interval $[0, T]$, where $T = 50$ min. We study four simulation cases, choosing the flux function (7) or the flux function (8):

Line	$\rho_{i,0}$	$\rho_{bi,0}$	Line	$\rho_{i,0}$	$\rho_{bi,0}$	Line	$\rho_{i,0}$	$\rho_{bi,0}$
1	0.4	0.4	21	0.3	/	41	0.1	/
2	0.35	0.35	22	0.2	/	42	0.1	/
3	0.3	/	23	0.1	/	43	0.25	0
4	0.2	/	24	0.1	/	44	0.3	/
5	0.35	0.35	25	0.2	/	45	0.4	0.4
6	0.2	0	26	0.1	/	46	0.3	0.3
7	0.25	/	27	0.2	/	47	0.2	/
8	0.4	0.4	28	0.25	/	48	0.4	0
9	0.35	0.35	29	0.2	0	49	0.35	/
10	0.3	/	30	0.4	/	50	0.3	0
11	0.2	/	31	0.35	0.35	51	0.2	/
12	0.1	/	32	0.3	0.3	52	0.1	0
13	0.1	/	33	0.2	/	53	0.1	/
14	0.25	/	34	0.35	/	54	0.2	0
15	0.3	/	35	0.2	/	55	0.1	/
16	0.4	0.4	36	0.25	/	56	0.2	0
17	0.3	0	37	0.4	/	57	0.25	/
18	0.2	/	38	0.35	/	58	0.2	0
19	0.4	0.4	39	0.3	/	59	0.15	0
20	0.35	0.35	40	0.2	/	60	0.15	0

Table 1. Initial conditions and boundary data for the lines of the network for case A.

Line	$\rho_{i,0}$	$\rho_{bi,0}$	Line	$\rho_{i,0}$	$\rho_{bi,0}$	Line	$\rho_{i,0}$	$\rho_{bi,0}$
1	0.4	0.4	19	0.4	0.4	48	0.5	0.7
2	0.5	0.5	20	0.5	0.5	50	0.5	0.7
5	0.5	0.5	29	0.4	0.7	52	0.4	0.7
6	0.4	0.7	31	0.4	0.4	54	0.5	0.7
8	0.4	0.4	32	0.4	0.4	56	0.4	0.7
9	0.5	0.5	43	0.4	0.7	58	0.5	0.7
16	0.4	0.4	45	0.4	0.4	59	0.5	0.7
17	0.4	0.7	46	0.5	0.5	60	0.5	0.7

Table 2. Initial conditions and boundary data for the lines of the network for case C.

1. at each node parameters, that optimize the cost functionals J_1 and J_2 (*optimal case*);
2. random α and q parameters (*static random case*) chosen in a random way at the beginning of the simulation process (for each simulation case, 100 static random simulations are made);
3. dynamic random parameters (*dynamic random case*) which change randomly at every step of the simulation process.

In the following pictures, we show the values of the functionals J_1 and J_2, computed on the whole network, as function of time. A legend for every picture indicates the different simulation cases.

The algorithm of optimization, which is of local type, can be applied to complex networks, without compromising the possibility of a global optimization. This situation is evident if we

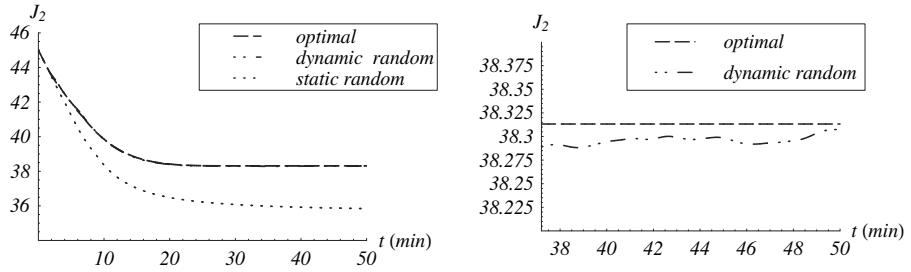


Fig. 10. J_1 for flux function (8), case A, and zoom around the optimal and dynamic random case (right).

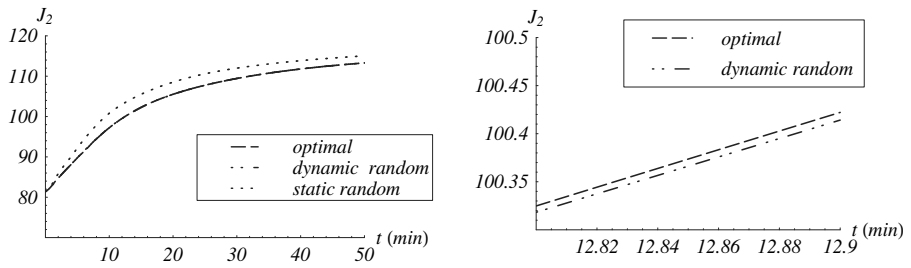


Fig. 11. J_2 for flux function (8), case B, and zoom around the optimal and dynamic random case (right).

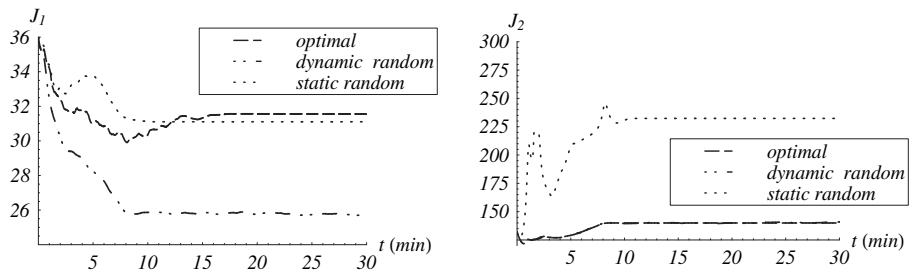


Fig. 12. J_1 and J_2 for flux function (7), case C.

consider the behaviour of J_1 for case A and J_2 for case B. For cases A and B, the cost functionals simulated with flux function (7) are constant, which is not surprising since the initial data on the lines is less than $\frac{1}{2}$. In case C, we present the behaviour of the cost functionals J_1 and J_2 for flux function (7). Boundary data are of Dirichlet type (unlike case A and B where we have considered Neumann boundary conditions) and the network is simulated with high incoming fluxes for the incoming lines and high initial conditions for inner lines. We can see, from Figure 12, that J_1 and J_2 are not constant as in cases A and B. Moreover, we have to take in mind that we have two different optimization algorithms for J_1 and J_2. Notice that the dynamic random case follows the optimal case for J_2 and not for J_1. Indeed, the optimal algorithm for J_1 presents an interesting aspect. When simulation begins, it is worst than the static random configuration. In the steady state, instead, the optimal configuration is the highest.

As for the dynamic random simulation, its behaviour looks very similar to the optimal one for cases A and B (for case C, only J_2 presents optimal and dynamic random configurations, that are very similar). Hence, we could ask if it is possible to avoid the optimization of the network, and operate in dynamic random conditions. Indeed, this last case originates strange phenomena, that cannot be modelled, hence it is preferred to avoid such a situation for telecommunication network design. To give a confirmation of this intuition, focus the attention on line 13, that is completely inside the network and it is strongly influence by the dynamics at various nodes. In Figure 13, we see that, using optimal parameters, the density on line 13 shows a smoother profile than the one obtained through a dynamic random simulation.

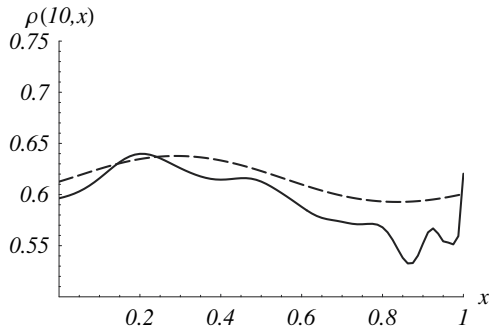


Fig. 13. Behaviour of the density on line 13 of the network of Figure 9, for $t = 10$, flux function (7), case C, in optimal and dynamic random simulations. Dashed line: optimal simulation for J_2; solid line: dynamic random simulation.

6. Conclusions

A fluid-dynamic model for data networks has been described. The main advantages of this approach, with respect to existing ones, can be summarized as follows. The fluid-dynamic models are completely evolutive, thus they are able to describe the traffic situation of a network every instant of time, overcoming the difficulties encountered by many static models. An accurate description of queues formation and evolution on the network is possible. The theory permits the development of efficient numerical schemes for very large networks. The model is based on packets conservation at intermediate time scales, whose flux is determined via a loss probability function (at fast time scales) and on a semilinear equation for the evolution of the percentage of packets going from an assigned source to a given destination.
The choice of the loss probability function is of paramount importance in order to achieve a feasible model. The fluid dynamic model has been compared with those obtained using various queueing paradigms, from proportional/excess to models with finite capacity, including different distributions for packet sizes. The final result is that such models give rise to velocity profiles and flux functions which are quite similar to the fluid dynamic ones.
In order to solve dynamics at node,Riemann Solvers have been defined considering different traffic distribution functions (which indicate for each junction J the outgoing direction of traffic that started at source s, has d as final destination and reached J from an assigned incoming road) and rules RA1 and RA2. The algorithm RA1, already used for road traffic models, requires the definition of a traffic distribution matrix, whose coefficients describe the percentage of packets, forwarded from incoming lines to outgoing ones. Using the algorithm

RA2, not considered for urban traffic as redirections are not expected from modelling point of view (except in particular cases, as strong congestions or road closures), priority parameters, indicating priorities among flows of incoming lines, and distribution coefficients have to be assigned.
The main differences between the two algorithms are the following. The first one simply sends each packet to the outgoing line which is naturally chosen according to the final packet destination. The algorithm is blind to possible overloads of some outgoing lines and, by some abuse of notation, is similar to the behaviour of a "switch". The second algorithm, on the contrary, sends packets to outgoing lines in order to maximize the flux both on incoming and outgoing lines, thus taking into account the loads and possibly redirecting packets. Again by some abuse of notation, this is similar to a "router" behaviour. Hence, RA1 forwards packets on outgoing lines without considering the congestion phenomena, unlike RA2. Observe that a routing algorithm of RA1 type working through a routing table, according to which flows are sent with prefixed probabilities to the outgoing links, is of "distance vector" type. Reverse, an algorithm of RA2 type can redirect packets on the basis of link congestions, so it works on the link states (hence on their congestions) and so it is of "link-state" type.
The performance analysis of the networks was made through the use of different cost functionals, measuring average velocity and average travelling time, using the model consisting of the conservation law. The optimization is over parameters, which assign priority among incoming lines and traffic distribution among outgoing lines. A complete solution is provided in a simple case, and then used as local optimal choice for a complex test network. Three different choices of parameters have been considered: locally optimal, static random, and dynamic random (changing in time). The local optimal outperforms the others. Then, the behaviour of packets densities on the lines, that permits to rule out the dynamic random case has been analyzed.
All the optimization results have been obtained using a decentralized approach, i.e. an approach which sets local optimal parameters for each junction of the network. The cooperative aspect of such decentralized approach is the following. When a router optimizes the (local) functionals, it takes into considerations entering and exiting lines. Such lines reach other nodes, which benefit from the optimal choice. This in fact reflects in good global behavior as showed by simulations, described below. In future we aim to extend the optimization results to more general junctions and to explore global optimization techniques.

7. References

Alderson, D.; Chang, H.; Roughan, M.; Uhlig, S. & Willinger, W. (2007). The many facets of internet topology and traffic, *Networks and Heterogenous Media*, Vol. 1, Issue 4, 569–600, ISSN 1556-1801.

Baccelli, F.; Chaintreau, A.; De Vleeschauwer, D. & McDonald, D. (2006). HTTP turbulence, *Networks and Heterogeneous Media*, Vol. 1, 1–40, ISSN 1556-1801.

Baccelli, F; Hong, D. & Liu, Z. (2001). Fixed points methods for the simulation of the sharing of a local loop by large number of interacting TCP connections, *Proceedings of the ITC Specialist Conference on Local Loop*, 1–27, Barcelona, Spain, (also available in Technical Report RR-4154, INRIA, Le Chesnay Cedex, France), ISBN 0249-6399.

Bretti, G.; Natalini, R. & Piccoli, B. (2006). Numerical approximations of a traffic flow model on networks, *Networks and Heterogeneous Media*, Vol. 1, 57–84, ISSN 1556-1801.

Bressan, A. (2000). *Hyperbolic Systems of Conservation Laws - The One-dimensional Cauchy Problem*, Oxford University Press, ISBN 0198507003, Oxford.

Cascone, A.; D'Apice, C.; Piccoli, B. & Raritá, L. (2007). Optimization of traffic on road networks, *Mathematical Models in Applied Sciences*, Vol. 17, 1587–1617, ISSN 0218-2025.

Cascone, A.; Marigo, A.; Piccoli, B. & Rarità, L. (2010). Decentralized optimal routing for packets flow on data networks, *Discrete and Continuous Dynamical Systems - Series B (DCDS - B)*, Vol. 13, No. 1, 59–78, ISSN 15313492.

Coclite, G.; Garavello, M. & and Piccoli, B. (2005). Traffic Flow on a Road Network, *SIAM Journal on Mathematical Analysis*, Vol. 36, 1862–1886, ISNN 0036-1410.

Dafermos, C. (1999). *Hyperbolic Conservation Laws in Continuum Physics*, Springer-Verlag, ISBN 354064914X, New York.

Daganzo, C. (1997). *Fundamentals of Transportation and Traffic Operations*, Pergamon-Elsevier, ISBN 0080427855, Oxford.

D'Apice, C.; Manzo, R. & Piccoli, B. (2006). Packet flow on telecommunication networks, *SIAM Journal on Mathematical Analysis*, Vol. 38, No. 3, 717–740, ISNN 0036-1410.

D'Apice, C.; Manzo, R. & Piccoli, B. (2008). A fluid dynamic model for telecommunication networks with sources and destinations, *SIAM Journal on Applied Mathematics (SIAP)*, Vol. 68, No. 4, 981–1003, ISSN 0036-1399.

D'Apice, C.; Manzo, R. & Piccoli, B. (2010). On the validity of fluid-dynamic models for data networks, *Journal of Networks*, submitted, ISSN 1796-2056.

Garavello, M. & Piccoli, B. (2006). *Traffic flow on networks*, AIMS Series on Applied Mathematics, vol. 1, American Institute of Mathematical Sciences, ISBN 1601330006, United States.

Godlewsky E. & Raviart, P. (1996). *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, Springer Verlag, ISBN 978-0-387-94529-3, Heidelberg.

Garavello, M. & Piccoli, B. (2005). Source-Destination Flow on a Road Network, *Communication in Mathematical Sciences*, Vol. 3, 261–283, ISSN 1539-6746.

Godunov, S. K. (1959). A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics, *Mat. Sb.*, Vol. 47, 271–306, ISSN 0368-8666.

Holden, H. & Risebro, N. H. (1995). A mathematical model of traffic flow on a network of unidirectional roads, *SIAM Journal on Mathematical Analysis*, Vol. 26, 999–1017, ISNN 0036-1410.

Marigo, A. (2006). Optimal distribution coefficients for telecommunication networks, *Networks and Heterogeneous Media*, Vol. 1, 315–336, ISNN 1556-1801.

Kelly, F.; Maulloo, A. K. & Tan, D. K. H. (1998). Rate control in communication networks: shadow prices, proportional fairness and stability, *Journal of the Operational Research Society*, Vol. 49, 237–252, ISSN 0160-5682.

Lighthill, M. J. & Whitham, G. B. (1955). On kinetic waves. II. Theory of Traffic Flows on Long Crowded Roads, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, Vol. 229, 317–345, doi: 10.1098/rspa.1955.0089.

Newell, G. F. (1980). *Traffic Flow on Transportation Networks*, MIT Press, ISBN 0262140322, Cambridge (MA,USA).

Olsén, J. (2003). On Packet Loss Rates used for TCP Network Modeling, *Technical Report*, Uppsala University.

Richards, P. I. (1956). Shock Waves on the Highway, *Oper. Res.*, Vol. 4, 42–51, ISSN 0030-364X.

Serre, D. (1999). *Systems of conservation laws I and II*, Cambridge University Press, ISBN 521582334, 521633303, Cambridge.

Tanenbaum, A. S. (2003). *Computer Networks*, Prentice Hall, ISBN 0130661023, Upper Saddle River.

Wierman, A.; Osogami, T. & Olsén, J. (2003). A Unified Framework for Modeling TCP-Vegas, TCP-SACK, and TCP-Reno, *Proceedings of the IEEE/ACM International Symposium on modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, 269–278, ISBN 0-7695-2039-1, Orlando, Florida, October 2003, Los Alamitos, California, Washington.

Willinger, W. & Paxson, V. (1998). Where Mathematics meets the Internet, *Notices of the AMS*, Vol. 45, 961–970, ISSN 0002-9920.

Edited by Jesús Hamilton Ortiz

This book guides readers through the basics of rapidly emerging networks to more advanced concepts and future expectations of Telecommunications Networks. It identifies and examines the most pressing research issues in Telecommunications and it contains chapters written by leading researchers, academics and industry professionals. Telecommunications Networks - Current Status and Future Trends covers surveys of recent publications that investigate key areas of interest such as: IMS, eTOM, 3G/4G, optimization problems, modeling, simulation, quality of service, etc. This book, that is suitable for both PhD and master students, is organized into six sections: New Generation Networks, Quality of Services, Sensor Networks, Telecommunications, Traffic Engineering and Routing.

ISBN 978-953-51-5611-6

Photo by PhonlamaiPhoto / iStock

IntechOpen